

Technische Universität Dresden

Entwurf und Modellierung von Multikanal-CMOS-Farbsensoren

Stephan Henker

der Fakultät Elektrotechnik und Informationstechnik
der Technischen Universität Dresden

zur Erlangung des akademischen Grades eines

Doktoringenieurs

(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender: Prof. Dr. G. Gerlach

Gutachter: Prof. Dr. R. Schüffny
Prof. Dr. W. Fischer
Prof. Dr. H. Beikirch

Tag der Einreichung 7.3.2005

Tag der Bestätigung: 1.8.2005

Abstract

Color image acquisition and image processing have become a key in modern data application. In order to provide high quality images, the field of accurate acquisition is most important in respect to all further processing steps. But a whole variety of current image sensors possess incorrect color rendition due to insufficient accuracy of optical sensor parameters. This is detrimental especially for color sensors, because in these cases specific color information will be incorrectly acquired. Further, traditional color correction methods do not use information on the specific sensor spectral sensitivity, thus losing substantial information for color correction.

The problem is investigated by introducing an algorithmic correction method which is capable of correcting dysfunctional sensor properties. The correction method is based on an enhancement of the CIE color perception model. According to this, color perception is modelled as a special integral transformation, where the spectral sensitivities of the photo receptors represent the base functions of the transformation. It is shown that different sets of photo receptors show the same perception, when their spectral sensitivities are linear dependent. On the other hand, photo receptors with no linear dependency show different perception and there is no analytical transformation between them. Thus, a perfect color correction is only possible if photo sensor and human perception show a linear dependency. In case of dissentient sensor characteristics, the correction method of spectral reconstruction can determine an optimal solution using a least square error optimization. Applying sensors with more than three color channels, this correction method can show improved results due to a better approximation.

For implementation of the color correction scheme, different sensor designs have been developed. Compared with currently dominating CCD (Charge Coupled Device) technology, a realisation of image sensors based on CMOS technology show a high potential. CMOS technology allow the integration of the sensor together with control and image processing on the same chip, thus enabling the design of sensor systems at low cost. But modern sub-100nm technologies show also substantial disadvantages, such as increased leakage currents. Special circuit designs have been developed to especially reduce the influence of leakage currents.

For application of the color correction method, new multi-channel photo sensors using vertically stacked photo diodes have been developed. The work further shows different concepts of multi-channel sensors capable of high quality color rendition. This approach is demonstrated on several new CMOS sensor designs with examples, implemented in a 90nm Infineon technology.

Einleitung

Digitale Farbbildkameras haben sich im letzten Jahrzehnt vom Nischenprodukt zum dominanten Bildaufnahmesystem entwickelt. Sie ersetzen dabei nicht nur den herkömmlichen Farbfilm, sondern eröffnen zunehmend auch neue Anwendungsgebiete. Die neuesten Entwicklungen auf dem Gebiet der elektronischen Bildaufnahme stellen Kameras auf CMOS-Basis dar. Gegenüber den momentan auf diesem Gebiet dominierenden CCDs (*Charged-Coupled Device*) bieten CMOS-Sensoren einige Vorteile, wie beispielsweise geringerer Verlustleistung und der Möglichkeit zur Integration verschiedener Systemkomponenten auf einem Chip.

Allgemein betrachtet, besteht die Aufgabe eines Farbbildaufnehmers darin, ein Bild aufzunehmen, damit es gespeichert und zu einem späteren Zeitpunkt wiedergegeben werden kann. Dabei ist üblicherweise impliziert, daß ein menschlicher Beobachter bei Betrachtung des wiedergegebenen Bildes die gleiche Farbwahrnehmung erfährt wie der Betrachter der originalen Szene. In der Praxis besteht allerdings das Problem, daß die Ausgaben eines technischen Bildsensors die gestellten Anforderungen nicht erfüllen. Es ist daher erforderlich, die Ausgaben des Sensors zu korrigieren, um eine farbrichtige Wiedergabe zu ermöglichen. Dieser Schritt wird allgemein als Farbkorrektur bezeichnet. Der wesentliche Gegenstand dieser Arbeit besteht in der Entwicklung und Implementierung von neuartigen Farbbildsensoren mit verbesserten Möglichkeiten zur Farbkorrektur. Die angewendeten Methoden und Parameter für die Implementierungen werden dabei allgemein aus Betrachtungen des menschlichen Sehvorgangs als biologischem Prozeß und aus bereits vorhandenen Standardisierungen des Farbsehens gewonnen.

Da auf dem Gebiet der Farbwahrnehmung jedoch viele verschiedene Modelle und Methoden existieren, die sich teilweise widersprechen, für eine umfassende Beschreibung zu stark vereinfacht sind oder gar auf unzutreffenden Annahmen beruhen, werden die notwendigen Grundlagen dieser Arbeit in einer einführende Beschreibung zusammengestellt. Diese geht von der menschlichen Farbwahrnehmung als biologischem Prozeß aus und führt über eine Erläuterung und Erweiterung der existierenden Standards zu den technischen Anwendungen.

In einem einführenden Kapitel wird daher auf die relevanten Teile des biologischen Farbwahrnehmungsprozesses eingegangen und ein zweites Kapitel widmet sich der technischen Beschreibung der menschlichen Farbwahrnehmung als Modell der biologischen Prozesse. Dabei werden die bestehenden Standardisierungen erweitert und in ein mathematisches Modell überführt, das eine verallgemeinerte Beschreibung der Farbaufnahme erlaubt.

Daraus folgernd soll die Notwendigkeit von Farbkorrekturen in technischen Bildaufnahmesystemen erläutert und mit Verweis auf existierende Ansätze eine allgemeine Methode zur Farbkorrektur abgeleitet werden. Die Anwendung dieser Methode und deren Möglichkeiten soll an verschiedenen Beispielen aus dem Bereich der Farbbildaufnahme illustriert werden.

Die abschließenden Kapitel widmen sich der technischen Umsetzung der gewonnenen Erkenntnisse und dem Entwurf von neuartigen elektronischen Farbbildsensoren. Die besonderen Herausforderungen liegen dabei im wesentlichen in der Entwicklung und Anwendung von neuen Entwurfsmethoden für Photosensoren in sub-100nm-CMOS-Technologien.

Inhaltsverzeichnis

Notationen und Formelzeichen	11
----------------------------------------	----

Kapitel I

Das menschliche Auge	13
1 Anatomie	15
1.1 Der Aufbau des Auges	15
1.2 Der Sehprozeß	16
2 Biologische Definition von Farbe	18
2.1 Die Farbwahrnehmung des Auges	18
2.2 Farbsinnstörungen und Farbenblindheit	19
2.3 Metamerie	20
2.4 Der Rezeptor-Farbraum	21
2.5 Farbverarbeitung im Auge	22
2.6 Farbadaptation und Umstimmung	23

Kapitel II

Modellierung des Farbsehens	25
1 Die Modellierung von Licht und Farbe	27
1.1 Die Definition von Helligkeit	27
1.2 Technische Definition von Farbe	27
1.3 Farbmessung und Farbmetriken	27
2 Die CIE-Farbdefinition	28
2.1 Vorbetrachtung	28
2.2 Farbmischversuche	29
2.2.1 Mathematische Betrachtung der Farbwahrnehmung	30
2.3 Die CIE Spektralwertfunktion	33
2.4 Das CIE-System	35
2.5 Nachteile des CIE-Systems	37
2.5.1 MacAdam-Schwellenellipsen	38
3 Reproduktion von Farbe	40
3.1 Additive Farbmischung	40
3.2 Subtraktive Farbmischung	41
3.3 Anwendung der additiven Farbmischung	43
3.4 Weißpunkt	45
3.5 Konstanz der Wahrnehmung	47
3.6 Klassifikation der Wiedergabe von Farbe	47

Kapitel III

Farbmessung und -korrektur	49
1 Farbmeßverfahren	51
1.1 Spektrale Farbmeßverfahren	52
1.2 Dreibereichsverfahren nach DIN5033	53
2 Farbkorrektur	53
2.1 Herkömmliche Methoden zur Farbkorrektur	54
3 Spektrale Rekonstruktion	56
3.1 Grundlagen	56
3.2 Die Methode der Spektralen Rekonstruktion	57

3.3	Vergleich mit nichtlinearen Farbkorrekturen	59
4	Anwendung der Spektralen Rekonstruktion	60
4.1	3-Kanal-Sensor	60
4.2	4-Kanal-Sensor	61
4.3	4-Kanal-CMOS-Sensor mit Dual-Junction-Photodioden	62
4.4	6-Kanal-CMOS-Sensor mit Triple-Junction-Photodioden	63
4.5	8-Kanal-CMOS-Sensor mit Dual-Junction-Photodioden	63
4.6	Schlußfolgerungen	64
5	Weißpunktkorrektur und Farbkonstanz	65
5.1	Weißpunktkorrektur	65
5.2	Farbkonstanz	66
5.3	Weißabgleich und Farbkonstanz mit Multikanal-Sensoren	67
5.4	Schlußfolgerungen	70

Kapitel IV

Licht- und Farbsensoren	73
1 Bildsensoren	75
1.1 Kenngrößen von Bildsensoren	75
1.2 Farbbildsensoren	76
1.2.1 Farbfiltermosaik	77
1.2.2 Multispektrale Farbbildaufnahme	78
2 Optische Filter für Farbbildsensoren	78
2.1 Physikalische Grundlagen	78
2.2 Interferenzfilter	79
2.3 Absorptionsfilter	80
3 Elektrische Verfahren zur Strahlungsmessung	81
3.1 Der äußere lichtelektrische Effekt	81
3.2 Vakuumphotozelle und Sekundärelektronenvervielfacher	81
3.3 Halbleiter-Photosensoren	82
3.3.1 Optische Generation	82
3.3.2 Photowiderstand	84
3.3.3 Photodiode	85
3.3.4 Photogate	87
3.3.5 Integrierender Photodetektor mit Photodiode	87
4 Pixelschaltungen auf CMOS-Basis	88
4.1 CCD und CMOS	88
4.2 Passive Pixel (PPS)	90
4.3 Aktive Pixel (APS)	91
4.3.1 Resetschaltungen	91
4.3.2 Einfache Ausleseschaltung	92
4.3.3 Ausleseschaltung mit elektronischer Shutter	93
4.4 Logarithmische Pixel	94
4.5 Digitale Pixel (DPS)	95
5 Rauschen in Photosensoren	96
5.1 Rauschterme	96
5.1.1 Thermisches Rauschen	97
5.1.2 kTC-Rauschen	98
5.1.3 Schrotrauschen	98
5.1.4 Photonenrauschen	99

5.1.5	1/f-Rauschen und Flicker-Rauschen	99
5.1.6	Quantisierungsrauschen	100
5.2	Rauschmodelle	100
5.2.1	Rauschmodell einer Photodiode	100
5.2.2	Rauschmodell eines MOS-Transistors	101
5.3	Rauschen von Pixelschaltungen	101
5.4	Methoden zur Rauschunterdrückung	102
5.4.1	Korrelierte Doppelabtastung	103
6	Fehler von Bildaufnahmesystemen	104
6.1	Allgemeine Betrachtungen	104
6.2	Alias-Effekte	105
6.3	Farbmoiré	106
6.3.1	Signaltheoretische Grundlagen	106
6.3.2	Modifikation des Korrekturansatzes	108

Kapitel V

	Beispielentwurf eines Multikanal-CMOS-Farbsensors	111
1	Grundlagen	113
1.1	Zielvorgaben	113
1.2	Multikanal-Photosensoren	114
1.3	Auswahl der Schaltungselemente	115
2	Entwurf von Multikanal-Sensoren in 0.35 µm-Technologie	117
2.1	Implementierung einer Pixelzelle	117
3	Entwurf von Multikanal-Sensoren in 90 nm-Technologie	118
3.1	Implementierungen von Pixelzellen	118
3.2	Operationsverstärker für geringe Betriebsspannungen	122
3.3	Funktionelle Erweiterungen der Pixelzellen	123
3.4	Konzeption einer Auslese/Korrekturschaltung	124
4	Meßergebnisse	126
	Zusammenfassung und Ausblick	129
	Literaturverzeichnis	131
	Eigene Publikationen als Hauptautor	139

Anhang A

	Physikalischer und mathematischer Anhang	141
1	Erläuterung physikalischer Größen	143
1.1	Sichtbares Licht	143
1.2	Radiometrische und Photometrische Grundgrößen	143
1.2.1	Bogenmaß , Raumwinkel	143
1.2.2	Strahlung, Lichtmenge	143
1.2.3	Strahlungsleistung, Lichtstrom	144
1.2.4	Strahlstärke, Lichtstärke	144
1.2.5	Bestrahlungsstärke, Beleuchtungsstärke	145
1.2.6	Photonenflußdichte	145
1.2.7	Belichtung	145
1.2.8	Strahldichte und Leuchtdichte	145

2	Mathematischer Anhang	146
2.1	Funktionsapproximation mit der Methode des kleinsten quadratischen Fehlers	146
3	Modellierungen	147
3.1	MOS-Transistormodell	147

Anhang B

Biologischer Anhang	149
1 Sehprozeß	151
1.1 Abbildung	151
1.2 Natürliche Abbildungsfehler	152
1.3 Transduktion	153
1.4 Hell-/Dunkeladaptation	155
1.5 Weber-Fechner-Gesetz	156
1.6 Signalauswertung und Weiterleitung in der Retina	157
1.7 Kontrast, Kontrastverstärkung	158
2 Kenngrößen des Auges	159
2.1 Sehschärfe	159
2.2 Ortsfrequenzabhängigkeit	161
2.3 Zeitliches Auflösungsvermögen	163

Anhang C

Wahrnehmungstheorie und Farbbeschreibung	165
1 Historie	167
1.1 Auge, Wahrnehmung und das Modell des Lichts	167
1.2 Physikalische Wahrnehmungs- und Farbtheorien	167
1.2.1 Von der Antike zur Neuzeit	167
1.2.2 Die Tristimulus-Theorie des Farbensehens	168
1.2.3 Die Young-Helmholtzsche Theorie	169
1.3 Phänomenologische Farbtheorien	170
1.3.1 Johann Wolfgang von Goethe	170
1.3.2 Die Duplizitätstheorie	171
1.3.3 Die Heringsche Gegenfarbentheorie	171
1.3.4 Modifikationen der Farbtheorien	173
1.4 CIE – Commission Internationale d'Eclairage	173
2 Farbbeschreibungssysteme	174
2.1 CIE XYZ und CIE xyY	174
2.2 Lineares/nichtlineares RGB und CMY/CMYK	174
2.3 CIE L^*a^*b	175
2.4 CIE L^*u^*v	176
2.5 CIE-LCH	178
2.6 Farbabstand DE	178
2.7 Gegenfarbsysteme	179
2.8 Farbsysteme und Farbordnungssysteme	179
Bemerkungen zu den japanischen Sprichworten	181

Notationen und Formelzeichen

- Physikalische Konstanten und Grundgrößen
 - E, W Energie, Arbeit
 - T absolute Temperatur
 - t, T_η Zeit als Variable, Zeit als Fixwert
 - f, ω, λ Frequenz, Kreisfrequenz und Wellenlänge
 - $c = 299'792'458 \text{ m/s}$ Lichtgeschwindigkeit
 - $k = 1.380'662 \cdot 10^{-23} \text{ J}\cdot\text{K}^{-1}$ Boltzmann-Konstante
 - $h = 6.626'176 \cdot 10^{-37} \text{ J}\cdot\text{s}$ Plancksches Wirkungsquantum
 - $e = 1.602'189 \cdot 10^{-19} \text{ A}\cdot\text{s}$ Elementarladung
 - $m_e = 9.109'534 \cdot 10^{-34} \text{ kg}$ Elektronenmasse
- Photometrische Größen
 - $Q, Q^\lambda, Q(\lambda)$ Strahlungsenergie, allgem., spektral und Lichtmenge
 - $\Phi, \Phi^\lambda, \Phi(\lambda)$ Strahlungsleistung, allgem., spektral und Lichtstrom
 - $\varphi(\lambda)$ spektrales Leistungsdichtespektrum
 - $I, I(\lambda)$ Strahlungsleistung pro Einheitsraumwinkel und Lichtstärke
 - $E, E^\lambda, E(\lambda)$ Bestrahlungsstärke, allgem., spektral und Beleuchtungsstärke
 - $\Pi^\lambda, \pi(\lambda)$ spektrale Photonenflußdichte und deren Funktion über λ
 - $L, L(\lambda)$ Strahldichte, Leuchtdichte
 - H Belichtung
 - $\alpha(\lambda)$ spektrale Absorptionsgrad
 - $\rho(\lambda)$ spektrale Reflexionsgrad/Reflektanz
 - $\tau(\lambda)$ spektrale Transmissionsgrad/Transmittanz
 - n_η Brechungsindex η
- Farbmétrische Größen
 - $\vartheta_\eta(\lambda)$ spektrale Empfindlichkeit des Farbrezeptors η
 - ε_η Erregung des Farbrezeptors η
 - $x(\lambda), y(\lambda), z(\lambda)$ Spektralwertfunktion des CIE-Normbeobachters
 - $\underline{\text{cie}}(\lambda) = [x(\lambda), y(\lambda), z(\lambda)]^T$ Vektor der CIE-Normspektralwertkurven
 - $\underline{b}_\eta(\lambda)$ Leistungsdichtespektrum der Primärfarbe η
 - \underline{B}_η Primärfarbsystem η
 - $\underline{b}_\eta(\lambda) = [b_1(\lambda), \dots, b_n(\lambda)]_\eta^T$ Vektor der Leistungsdichtespektren des Primärfarbsystems η
 - \underline{m} Farbvektor, für $|\underline{m}| = 3$: Dreifarbwert
 - $\underline{m}_{\text{cie}} = [X, Y, Z]^T$ Normfarbvektor
 - X, Y, Z Normfarbwerte
 - x, y, z Farbwertanteile
 - $\underline{a}(\lambda)$ spektrale Sensitivität, Spektralwertfunktion
 - $\varphi(\lambda)$ Funktion von Remissionswerten
 - g_η, k_η Faktor/Gewicht η
 - $u(\lambda)$ Spektrum des Umgebungslichts
 - $d65(\lambda)$ spektraler Verlauf des Normlichts D65

$\underline{s}(\lambda) = [s_1(\lambda), \dots, s_n(\lambda)]^T$	Vektor der spektralen Sensitivitäten eines praktischen Sensors
S_η	Ausgabe η eines praktischen Sensors
\underline{S}	Vektor der Ausgaben eines praktischen Sensors
$\underline{a}(\lambda) = [a_1(\lambda), \dots, a_n(\lambda)]^T$	Vektor der spektralen Sensitivitäten eines idealen Sensors
A_η	Ausgabe η eines idealen Sensors
\underline{A}	Vektor der Ausgaben eines idealen Sensors
$c_{i,j}$	Koeffizient i, j
$\mathbf{C}, \mathbf{Z}, \mathbf{X}$	Matrizen
\mathbf{E}	Einheitsmatrix

- Elektrische Größen

E, B	elektrische Feldstärke, magnetische Induktion
Q, C	Ladung, Kapazität
ε_η	Dielektrizitätszahl/konstante
$\beta(\lambda)$	Absorptionskoeffizient als Funktion der Wellenlänge
$G(\lambda, x)$	optische Generationsrate
τ_n, τ_t	Ladungsträgerlebensdauer und -transitzeit
$\eta_q(\lambda)$	Quantenwirkungsgrad als Funktion der Wellenlänge
$j_{ph}(\lambda)$	Photostromdichte als Funktion der Wellenlänge
$A(\omega)$	Übertragungsfunktion
B	Bandbreite
U_G	Spannung an Gate bezüglich Bulk-Potential
U_D	Spannung an Drain bezüglich Bulk-Potential
U_S	Spannung an Source bezüglich Bulk-Potential
β	Transistorkonstante
μ	Beweglichkeit der Ladungsträger
$C_{ox'}$	flächenbezogenen Oxidkapazität
W_{eff}	effektive Transistorweite
L_{eff}	effektive Transistorweite
n	Steilheitskonstante
V_{T0}	Schwellspannung
V_{DD}	Betriebsspannung
I_0	Sättigungsstrom
U_T	Temperaturspannung
I_D	Drainstrom

Kapitel I

Das menschliche Auge

朱に交われば赤くなる。

1 Anatomie

Das beste bisher bekannte Bildverarbeitungs- und Mustererkennungssystem ist das menschliche Auge in Verbindung mit der Bildauswertung durch das Gehirn. Der Sehvorgang selbst ist einer der am höchsten entwickelten biologischen Informationsverarbeitungsprozesse. Viele technische Systeme orientieren sich am Aufbau und der Funktionsweise des Auges. Als Fernziel von Bildverarbeitungssystemen wird dabei die Leistung des biologischen Vorbildes anvisiert. Zum besseren Verständnis der Bildverarbeitung soll daher zuerst auf das Auge, den Aufbau und die prinzipielle Funktion eingegangen werden.

1.1 Der Aufbau des Auges

Das Auge ist ein komplexes Sinnesorgan, das aus vielen verschiedenen, miteinander interagierenden Teilen besteht. Die für den Sehprozeß wichtigen Komponenten sind dabei die Hornhaut, die Regenbogenhaut (Iris), die Linse, der Glaskörper und die Netzhaut mit den lichtempfindlichen Rezeptorzellen. Bei landlebenden Tieren besitzt die Hornhaut die höchste Brechkraft am lichtbrechenden Apparat des Auges. Nachdem einfallendes Licht die Hornhaut passiert hat, trifft es auf Regenbogenhaut, Linse und Glaskörper. Die Iris dient dabei der Regulierung des Lichteinfalls. Dazu existiert eine bewegliche Öffnung in der Iris, die Pupille, die sich wie eine Lochblende verengen oder erweitern kann. Die Linse leistet zwar nur einen relativ geringen Beitrag zur Lichtbrechung, ist aber für die scharfe Abbildung auf der Netzhaut verantwortlich. Die als Akkommodation bezeichnete Anpassung des Auges an wechselnde Objektentfernungen erfolgt dabei durch Änderung der Linsenform. Die Netzhaut, Retina, besteht aus mehreren Zellschichten, die neben den lichtempfindlichen Rezeptorzellen noch Schichten aus vorverarbeitenden Horizontalzellen, Bipolarzellen, Amakrinzellen und Ganglienzellen enthält.

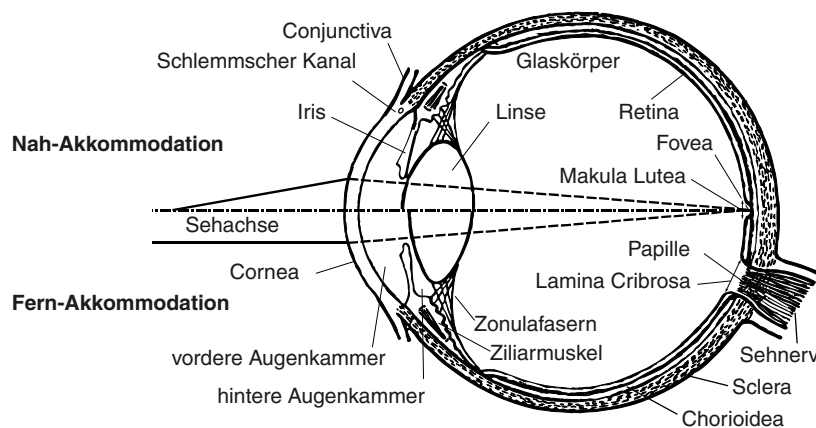


Bild I.1 Schematische Darstellung des menschlichen Auges

Treffen Lichtstrahlen, die von einem Gegenstand ausgehen, auf das Auge, so werden sie von der Hornhaut, der Linse und dem Glaskörper gebrochen, und ergeben ein umgekehrtes, reelles Bild auf der Netzhaut, das von den lichtempfindlichen Sinneszellen aufgenommen wird. Bei allen gut sehenden Wirbeltieren findet sich eine deutlich abgesetzte Partie in der Netzhaut (Area centralis) mit dicht angeordneten Sehzellen. Dadurch wird an dieser Stelle eine höhere Auflösung ermöglicht. Entwicklungsbedingt fallen die Lichtstrahlen nach dem Passieren von Linse und Glaskörper zuerst auf Schichten mit verarbeitenden Nervenzellen, dann auf die Zellkörper der in der untersten Schicht der Netzhaut liegenden Rezeptorzellen und erst zuletzt auf deren

lichtempfindliche Außenglieder. Dadurch wird zwangsläufig eine gewisse Unschärfe verursacht. An der Area centralis wird dies durch eine Grube (Fovea centralis) kompensiert, in der die Ganglienzellen und Nervenfasern zur Seite geschoben und schräg gelagert sind, so daß die Lichtstrahlen fast unmittelbar aus dem Glaskörper auf die Sehzellen fallen. Diese Stelle schärfsten Sehens wird auch als gelber Fleck (Macula lutea) bezeichnet und stellt den funktionell wichtigsten Teil der Netzhaut dar. Wird ein Objekt fixiert, werden Kopf und Augen automatisch so ausgerichtet, daß sich die Objekte des Interesses auf dem gelben Fleck abbilden.

Die Photorezeptoren, die durch die Aufnahme von Licht- und Farbreizen das Sehen ermöglichen, unterteilen sich in Zapfen und Stäbchen. Pro Auge besitzt der Mensch ca. 6,5 Millionen Zapfen. Diese sind hauptsächlich in und um die Fovea verteilt. Beim Menschen existieren drei verschiedene Typen von Zapfen, die sich in ihrer Sensitivität über den Wellenlängenbereich des sichtbaren Lichtes unterscheiden. Dadurch wird das Erkennen von Farben ermöglicht. Die Zapfen sind jedoch weniger lichtempfindlich als die Stäbchen und werden erst bei größeren Beleuchtungsstärken angeregt. Sie sind folglich für das Farbsehen am Tage, das phototopische Sehen, verantwortlich. Die 75-150 Millionen Stäbchen des Menschen sind wesentlich lichtempfindlicher, jedoch nur bei geringen Beleuchtungsstärken aktiv und bestimmen das menschliche Dämmerungs- und Nachtsehen. Bei größerer Lichtintensität werden sie übersättigt und tragen nicht mehr zum Sehvorgang bei. Stäbchen sind über die ganze Retina verteilt, mit Ausnahme der Fovea, in der keine Stäbchen vorhanden sind. Stäbchen besitzen alle dieselbe spektrale Sensitivität, so daß die verschiedenen Wellenlängen des eintreffenden Lichtes nicht unterschieden werden können. Bei geringen Lichtstärken ist somit keine Farbsicht möglich. Die Stäbchen dienen demnach dem farblosen Sehen bei geringen Lichtstärken, dem sogenannten skototopischen Sehen [MBM81].

1.2 Der Sehprozeß

Fällt Licht auf eine Rezeptorzelle, so wird es vom sogenannten Sehfärbstoff absorbiert und löst eine biochemische Reaktion aus. Der Lichtreiz wird dabei in eine elektrische Erregung gewandelt und noch in der Netzhaut vorverarbeitet. Der Vorgang der Absorption eines Lichtquants bis zur Erzeugung einer Potentialänderung eines Rezeptors, der Hyperpolarisation, wird als Transduktion bezeichnet. Dieser Prozeß ist durch eine örtliche und zeitliche Integration der von den absorbierten Lichtquanten verursachten Zellerregungen gekennzeichnet. Noch in den Rezeptorzellen erfolgt eine Bewertung des aufgenommenen Lichtreizes mit einer monotonen Funktion, die einen Schwellwert und eine Sättigung besitzt.

Eine Rezeptorzelle gibt ihre Information danach an mehrere Horizontal- sowie Bipolarzellen weiter, die eine Vorverarbeitung vornehmen. Über Amakrinzellen gelangen die Rezeptorsignale an die Ganglienzellen, die eine Umwandlung der vorverarbeiteten Signale von einer Amplitudenmodulation (Zellpolarisierung) in eine Impulsmodulation (neuronale Entladungsrate) vornehmen. Dabei erhält eine Ganglienzelle die vorverarbeitete Information von mehreren Horizontal-, Bipolar- und Amakrinzellen und somit von einer Gruppe von Rezeptoren, die das sogenannte rezeptive Feld des Neurons bilden [Har38]. Diese Zusammenfassung verschiedener Rezeptorerregungen führt zu einer stärkeren Aktivierung der zugehörigen Ganglienzelle und damit zu einer größeren Empfindlichkeit. Weiterhin überlappen sich die rezeptiven Felder mehrerer Ganglienzellen gegenseitig und besitzen unterschiedliche Größen. In der Fovea sind die

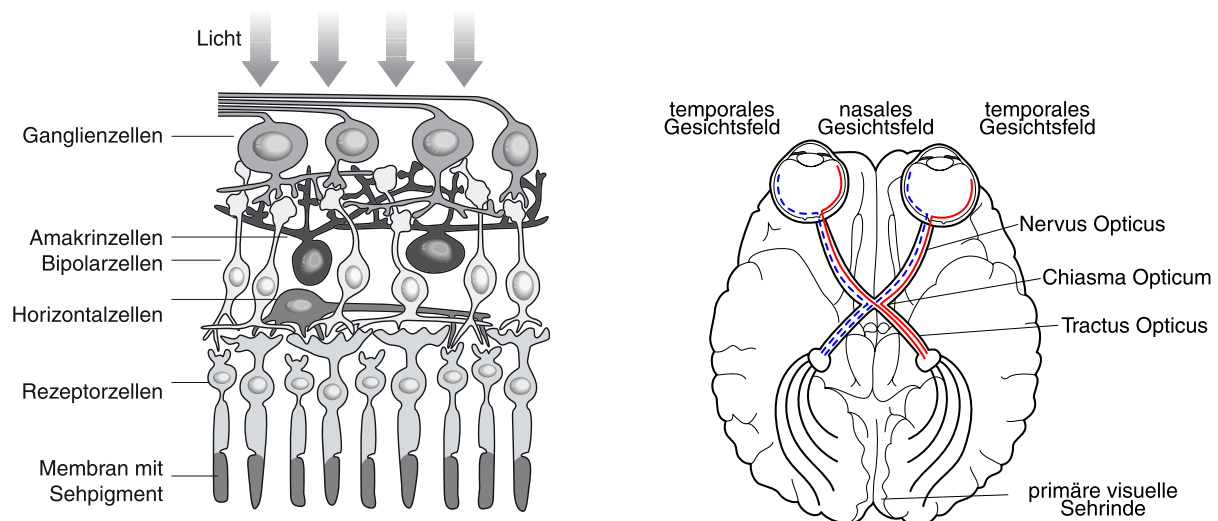


Bild I.2 Schematischer Aufbau der Retina und Verlauf des Sehnervs im Gehirn

Felder klein und bewirken damit ein verbessertes Auflösungsvermögen; zur Peripherie hin werden die rezeptiven Felder ausgedehnter. Weiterhin existiert ein Mechanismus zur Adaptation des Auges an die Lichtstärke durch Variation des Ausmaßes der rezeptiven Felder.

Die ca. 1.2 Millionen Ganglienzellen verlaufen von den Rezeptorzellen aus gesehen innen im Augapfel entlang und treffen sich im Sehnervenkopf. Dort treten die Fortsätze der Ganglienzellen durch eine runde, siebartige Öffnung in der Lederhaut in Bündeln aus dem Augapfel und vereinen sich zum Sehnerv (Nervus Opticus). An diese Austrittsstelle des Sehnervs aus dem Augapfel sind keine Rezeptorzellen ausgebildet, weshalb diese Stelle nicht lichtempfindlich ist. Sie wird als der blinde Fleck bezeichnet und befindet sich etwa 15 Grad exzentrisch auf der nasalen Seite der Retina. Diese Unterbrechung der Netzhaut ist nicht direkt wahrnehmbar, da sie vom Gehirn künstlich ausgefüllt wird.

Der Sehnerv verläuft 25 bis 40 Millimeter innerhalb der Augenhöhle und 10 bis 15 Millimeter innerhalb des Schädels, bevor sich die Sehnerven der beiden Augen vereinen und kreuzen, um dann ins Gehirn einzumünden (Bild I.2). Die Überkreuzung der Sehnerven ist dabei besonders markant. Dadurch wird der rechte Teil des Gesichtsfeldes, des Bildes, das die ruhenden Augen von der Außenwelt erfassen, auf das linke Sehfeld abgebildet. Der visuelle Kortex der linken Hirnhälfte empfängt somit Informationen der beiden linken Netzhauthälften, die ihrerseits die rechten Hälften des Gesichtsfeldes erfassen. Entsprechend übernimmt visuelle Kortex der rechten Hirnhälfte die linken Hälften des Gesichtsfeldes. Durch diese Überkreuzungen und das Vermischen der Information beider Augen ist räumliches Sehen möglich. Ein großer Teil der Sehnerven zieht nach der Überkreuzung ohne weitere Unterbrechung zu den sogenannten den äußeren Kniehöckern, zwei Zellanhäufungen im Gehirn, in denen wahrscheinlich die räumlichen Unterschiede registriert werden. Die Zellen der Kniehöcker sind durch ihre Nervenfasern unmittelbar mit dem primären Sehfeld verbunden. Vom primären Sehfeld aus werden die Informationen zu benachbarten Rindengebieten und an tiefer im Gehirn gelegene Stellen weitergeleitet. Einige Nervenfasern ziehen vom primären Sehfeld wieder zurück zu den Kniehöckern. Die Funktion dieser Rückkopplung ist bisher unbekannt. Das primäre Sehfeld stellt somit keineswegs das Ende, sondern nur eine Stufe in der Verarbeitung der visuellen Information dar. In den einzelnen Stufen von der Netzhaut bis zum Sehfeld des Gehirns erfolgen unterschiedliche

Verarbeitungs- und Abstraktionsschritte. So reagieren bestimmte Nervenzellen im Sehfeld der Hirnrinde vorzugsweise auf die Bewegung des Bildes auf der Netzhaut in einer bestimmten Richtung. Experimente haben weiterhin gezeigt, daß nur Informationen über Lichtänderungen an Grenzlinien an das Gehirn gesendet werden. Bereiche ohne Änderungen werden vom Gehirn gleichförmig ergänzt. Um solche Änderungen an Grenzlinien entdecken zu können, führen die Augen kleine Zitterbewegungen aus, sogenannte sakkadische¹ Augenbewegung, die das Bild des Gegenstandes auf der Netzhaut hin und her wandern lassen [MBM81]. Eine weitergehende Beschreibung von Funktionen und Kenngrößen des Auges ist in Anhang B zu finden.

2 Biologische Definition von Farbe

2.1 Die Farbwahrnehmung des Auges

Von den beiden Photorezeptortypen, den Stäbchen und Zapfen, sind die Zapfen zwar weniger lichtempfindlich, ermöglichen jedoch durch die unterschiedliche spektrale Sensitivität, das wahrgenommene Licht nicht nur nach der Intensität als Helligkeit, sondern auch nach der Wellenlänge in Form eines Farbeindrucks zu bewerten. Aus der Bandbreite der elektromagnetischen Wellen können Menschen allerdings nur den Bereich von ca. 400 bis 700 nm wahrnehmen, nicht jedoch die benachbarten Bereiche Ultraviolett und Infrarot.

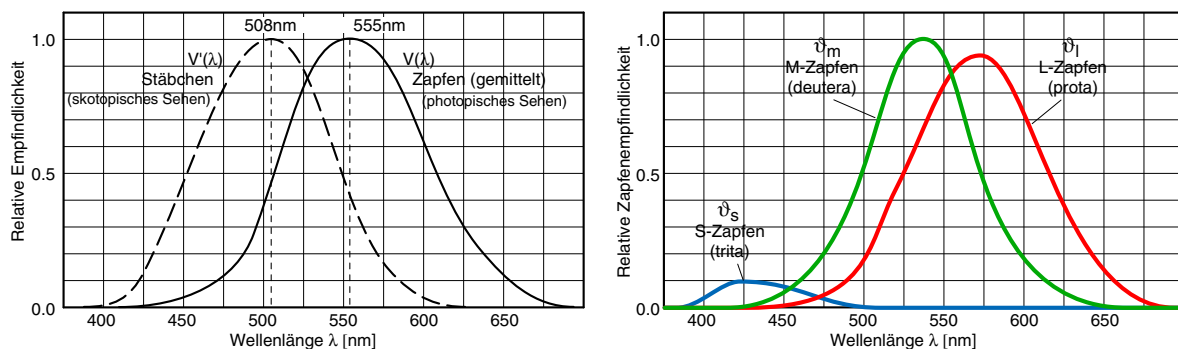


Bild I.3 Spektrale Empfindlichkeiten von Stäbchen und Zapfen beim Menschen

Allgemein existieren beim Menschen drei Arten von Zapfen. Nach dem Spektralbereich ihrer höchsten Empfindlichkeit werden sie lang- (L - *long*), mittel- (M - *middle*) und kurz- (S - *short*) wellenlängen-sensitiv genannt. Die medizinische Benennung nutzt die altgriechische Zählung *protia* (p - eins), *deutera* (d - zwei) und *trita* (t - drei). Die Empfindlichkeitsmaxima der Sinneszellen liegen für L-Zapfen bei 564 nm (Absorptionsbereich: ca. 430-720 nm), für M-Zapfen bei 534 nm (Absorptionsbereich: ca. 400-630 nm) und für S-Zapfen bei 419 nm (Absorptionsbereich: ca. 380-530 nm) (Bild I.3). Das Maximum der durchschnittlichen Empfindlichkeit liegt bei 555 nm [DBM83]. Im Vergleich dazu absorbieren die Stäbchen Licht im Wellenlängenbereich des sichtbaren Spektrums mit einem Absorptionsmaximum bei 508 nm (Absorptionsbereich: ca. 380-650 nm) (Bild I.3). Bei skotopischem Sehen (Nachtsicht) liegt die relative Empfindlichkeit somit bei kleineren Wellenlängen. Dieses Phänomen wird als Purkinje-Verschiebung² bezeichnet und führt u.a. dazu, daß im Dunkeln blaue Objekte heller wirken als rote, während es im Hellen umgekehrt ist.

1. frz. *saccade* = Ruck

2. nach Johannes Evangelista Purkinje (1787-1869)

Als Bezeichnung für die drei Spektralregionen mit der höchsten Empfindlichkeit der drei Zapfentypen wird hier allgemein die Bezeichnungen Blau, Grün und Rot verwendet. Dies ist nicht in allen Publikationen gleich gehalten. In einigen ist Violett, Grün und Orangerot im gleichen Sinne angewandt. In älteren Texten wird vielfach zwischen additivem und subtraktivem Blau (Blau und Cyan) sowie additivem und subtraktivem Rot (Rot und Magenta) unterschieden, während Grün und Gelb generell als getrennte Farbbezeichnungen geführt werden. Cyan wird auch als Blaugrün bezeichnet, Magenta als Purpur. Aufgrund der vielfältigen Bezeichnungen, die auch andere Farben betreffen und sich in verschiedenen Kontexten zuweilen auch widersprechen, ist eine allgemeine begriffliche Definition von Farben nur eingeschränkt möglich.

2.2 Farbsinnstörungen und Farbenblindheit

Durch individuelle genetische Variationen der Nervenzellen der Retina sind auch individuell verschiedene Reaktionen auf Lichtreize möglich. Am deutlichsten wird dies bei genetischen Defekten. Fehler in den Stäbchen führen dabei zur Störung des skotopischen Sehens und zu Nachtblindheit. Bei Einschränkungen von Zapfen können Farben nicht korrekt erkannt werden. Es existieren dabei verschiedene Arten und Ausmaße dieser sogenannten Farbsinnstörungen. Prinzipiell wird in Farbschwachsichtigkeit (Farbanomalie) und Farbenblindheit (Farbanopie) für verschiedene Farben eingeteilt. Dabei wird zwischen dem Mangel am rot-empfindlichen Rezeptor (protanomal, protanop), grün-empfindlichen Rezeptor (deutanomal, deutanop) und blau-empfindlichen Rezeptor (tritanomal, tritanop) unterschieden. Die häufigsten Farbsinnstörungen sind angeboren und werden aufgrund ihrer X-chromosomalen Kodierung geschlechtsgebunden vererbt, so daß deutlich mehr Männer (etwa 8% aller Männer) als Frauen (etwa 0,4% der Frauen) betroffen sind. Am häufigsten tritt eine Grünschwäche auf (4,2%), gefolgt von Rot-schwäche (1,6%), Grünblindheit (1,5%) und Rotblindheit (0,7%). Störungen im Blaubereich sind dagegen sehr selten ($10^{-3}\%$) und nicht geschlechtsspezifisch. Das zugehörige Gen befindet sich beim Menschen auf Chromosom 7. Noch seltener tritt eine totale Farbenblindheit (Stäbchenmonochromasie) auf, bei der lediglich Helligkeitsunterschiede wahrgenommen werden können ($10^{-4}\%$) [MWW]. Verschiedene Untersuchungen bei Menschen und Makaken haben ergeben, daß kein einheitliches Zapfenverteilungsbild existiert. Die Zapfen sind auf der Netzhaut in einem unregelmäßigen Mosaik angeordnet. Die Dichte ist in der Fovea am höchsten und nimmt zur Peripherie hin ab. In der zentralen Fovea befinden sich dabei ausschließlich L- und M-Zapfen. Auch bei farbtüchtigen Personen befinden sich dort keine Rezeptoren für kurze Wellenlängen. Beim normalsichtigen Menschen sind nur ungefähr 7% der Farbzeptoren S-Zapfen und ausschließlich in der Peripherie der Netzhaut zu finden. Weiterhin sind beim Menschen L-Zapfen tendenziell stärker vertreten als M-Zapfen. Die publizierten Verhältnisse variieren von 1:1 bis 3:1. Bei Makaken sind L- und M-Zapfen ungefähr gleich häufig. Die individuelle Variabilität dieses Zahlenverhältnisses ist dabei sowohl bei Makaken als auch bei Menschen sehr hoch [Lee96]. Es ist bisher unbekannt, wodurch die retinale Zapfenverteilung bestimmt wird. Die Auswirkungen der unterschiedlichen L- und M-Zapfenanteile der Retina auf die Wahrnehmung ist ebenfalls nur wenig untersucht. Extreme Verhältnisse liegen jedoch bei Rot-Grün-Blinden vor, die entweder keine L- oder keine M-Zapfen besitzen. Es ist daher zu vermuten, daß auch bei leichten Differenzen des L- und M-Zapfenverhältnisses Wahrnehmungsunterschiede zwischen den Individuen bestehen.

2.3 Metamerie

Ausgehend vom biologischen Aufbau des Auges wird die Empfindung einer spezifischen Farbe durch die Reizung der drei Zapfentypen vollständig bestimmt. Dabei existieren verschiedene Spektren von Lichtquellen oder reflektierenden Gegenständen, die beim Menschen zu einem identischen Farbeindruck führen. Beispielsweise läßt sich aus reinen Spektralfarben im roten, grünen und blauen Bereich weißes Licht mischen. Das Spektrum der Mischung besitzt drei einzelne Spektrallinien. Dieses Linienspektrum führt zu der identischen Wahrnehmung „Weiß“ wie ein kontinuierliches Spektrum, das alle Wellenlängen zu gleichen Teilen enthält. Solche spektrale Intensitätsverteilungen, die trotz unterschiedlicher Zusammensetzung die gleiche Farbwahrnehmung hervorrufen, heißen bedingt gleich bzw. metamer. Lichter mit gleichen Spektren sind dagegen isomer.

Das Konzept von Metamerie ist ein wichtiger Aspekt in der praktischen Anwendung von Farbe. Alle Farbreproduktionen basieren auf der Möglichkeit, metamere Farben mit unterschiedlichen Spektren zu erzeugen. Bei jedem farbig reproduzierten Bild sind die Spektren, die vom ursprünglichen Gegenstand hervorgerufen wurden, idealerweise metamer zu den Spektren des Wiedergabemediums (Druck, Bildschirm etc.). Auch in der Fotografie wird versucht, Farben so zu reproduzieren, wie sie vom menschlichen Beobachter gesehen werden. Dieser Prozeß ist jedoch nicht umkehrbar. Die Analyse der spektralen Zusammensetzung von Licht ist auf Basis einer Farbdarstellung nicht möglich. So kann beispielsweise das Spektrum der Abstrahlung eines Gegenstandes nicht aus einer Farbfotografie ermittelt werden. Weiterhin sind für den Menschen metamere Farben nicht unbedingt auch für andere Lebewesen metamer und umgekehrt. Aufgrund der individuellen Unterschiede in den Zapfenverhältnissen und deren Verteilung auf der Netzhaut ist zudem eine differenzierte Wahrnehmung von metameren Spektren auch zwischen Menschen wahrscheinlich. Die Wahrnehmung von Farben ist damit als subjektive Empfindung nur eingeschränkt allgemein charakterisierter.

Die Ursache der Metamerie liegt im Funktionsprinzip der Photorezeption des menschlichen Sehsystems mit den drei verschiedenen Zapfentypen begründet. Wird das Gesetz der spezifischen Sinnesenergie von Johannes Müller³ auf das Auge übertragen, folgt daraus, daß zwei beliebige Lichtreize, die vom Sehfarbstoff (Rhodopsin bzw. Jodopsin) gleich absorbiert werden, auch zu einer gleichen Reaktion der Zapfen führen (Prinzip der Univarianz [Rush72]). Die spektrale Empfindlichkeit eines einzelnen Zapfentyps beschreibt dabei die Wahrscheinlichkeit, mit der ein Photon absorbiert wird, als Funktion seiner Energie bzw. der Wellenlänge. Die resultierende Zapfenerregung hängt dann nur noch von der Anzahl der absorbierten Photonen ab, nicht aber von deren Wellenlänge. Daraus folgt, daß die spektrale Empfindlichkeit der Zapfentypen proportional zu deren Absorptionsspektrum ist. Wyszecki und Stiles [WS82] zeigen, daß tatsächlich ein linearer Zusammenhang zwischen den spektralen Absorptionsfunktionen der drei Farbrezeptortypen und den empirisch ermittelten Farbempfindungskurven besteht, die in Kapitel II, Abschnitt 2 vorgestellt werden.

3. Johannes Müller (1801-1858) formulierte 1826 das Gesetz der spezifischen Sinnesenergie, gemäß dem gleiche Sinnesreize auch gleiche Empfindungen auslösen.

2.4 Der Rezeptor-Farbraum

Prinzipiell sind die Erregungen ε_l , ε_m und ε_s der Photorezeptoren des Auges für den lang-, mittel und kurzwelligen Bereich wie folgt bestimmt:

$$\varepsilon_{l, m, s} = \int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot \vartheta_{l, m, s}(\lambda) \cdot d\lambda \quad (\text{I.1})$$

mit dem Leistungsdichtespektrum des einfallenden Lichtes $\varphi(\lambda)$ und der Empfindlichkeit der einzelnen Rezeptoren ϑ_l , ϑ_m und ϑ_s entsprechend Bild I.3. In Vektorschreibweise ergibt sich:

$$\underline{\varepsilon} = \int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot \underline{\vartheta}(\lambda) \cdot d\lambda \quad (\text{I.2})$$

Dabei stellt $\underline{\varepsilon} = [\varepsilon_l, \varepsilon_m, \varepsilon_s]^T$ den Erregungsvektor der Rezeptoren dar. Der Vektor der spektralen Empfindlichkeiten ist $\underline{\vartheta}(\lambda) = [\vartheta_l(\lambda), \vartheta_m(\lambda), \vartheta_s(\lambda)]^T$. Da die biologischen Rezeptorzellen jedoch Schwelle und Sättigung aufweisen, die zudem über Adaptationsmechanismen von der mittleren Helligkeit abhängen, ist das Ausgangssignal der Zelle eine monotone Funktion der Erregung. In der physikalischen Betrachtung bildet die Menge aller möglichen Lichtspektren $\varphi(\lambda)$ einen unendlich-dimensionalen Raum, da die Intensität für jede Wellenlänge unabhängig variieren kann. Da jedoch keine Spektralfunktionen mit negativen Intensitäten existieren, ist dieser Raum kein Vektorraum im Sinne der linearen Algebra. Wird $\varphi'(\lambda) = g \cdot \delta(\lambda - \lambda_0)$ definiert, wobei δ den Dirac-Impuls darstellt, ergibt sich $\varphi'(\lambda)$ als Linienspektrum mit einer einzigen Spektrallinie der Stärke g an der Stelle λ_0 . Diese reinen Spektralfarben bilden ein geeignetes Koordinatensystem des Spektralraumes. Durch die drei verschiedenen Photorezeptoren des menschlichen Auges wird der unendlich-dimensionale Raum der Spektren auf den dreidimensionalen Raum der Rezeptorerregungen $\underline{\varepsilon} = [\varepsilon_l, \varepsilon_m, \varepsilon_s]^T$ projiziert. Der Rezeptorraum ist dabei ein linearer Unterraum des Spektralraumes, wobei jedoch die Einschränkung auf positive Intensitäten der Spektren und positive Erregungen existiert. Mit dieser Dimensionsreduzierung ist ein Informationsverlust verbunden, der die eigentliche Ursache der Metamerie darstellt. Dabei existieren unendlich viele verschiedene Spektren, die auf einen einzigen Erregungsvektor abgebildet werden und damit trotz unterschiedlicher spektraler Leistungsdichten eine identische Farbwahrnehmung bewirken. Geht man weiter davon aus, daß jeder Zapfentyp Erregungen im Intervall $[0, 1]$ annehmen kann, so läßt sich der Raum der Rezeptorerregungen als Würfel $[0, 1]^3$ darstellen. Tatsächlich existieren jedoch nicht zu allen Vektoren dieses Rezeptorraumes zugehörige Spektren. So gibt es auf Grund der Überlappungen der Absorptionskurven kein physikalisches Spektrum, um einen Rezeptor einzeln zu reizen. Erregungsvektoren der Art $\underline{\varepsilon} = [1, 0, 0]^T$ können demnach nicht auftreten. Der Rezeptorraum als Menge aller möglichen Erregungsvektoren läßt sich mit folgender Überlegung bestimmen: Zunächst sind die Erregungsvektoren, die durch reine Spektralfarben erzeugt werden, im Rezeptorraum enthalten: $\underline{\varepsilon}(\lambda_0) = [\vartheta_l(\lambda_0), \vartheta_m(\lambda_0), \vartheta_s(\lambda_0)]^T$. Diese Erregungen bilden eine Linie im Raum der Rezeptorerregungen, die Spektralfarbenzug genannt wird. Weiterhin ist $\underline{\varepsilon} = [0, 0, 0]^T$ Element des Erregungsraumes. Alle weiteren Farben sind Linearkombinationen der genannten Vektoren mit positiven Koeffizienten. Formal ergibt sich der Rezeptorraum somit als konvexe Hülle⁴ der Vektoren des Spektralfarbenzug (Bild I.4).

Aus den Zapfenempfindlichkeiten ergibt sich, daß die Menge der möglichen Rezeptorerregun-

4. Die konvexe Hülle einer Menge M ist die kleinste Menge N , die M enthält und dazu alle Punkte x mit der Eigenschaft: $x = m \cdot a + (1 - m) \cdot b$ für zwei beliebige Elemente $a, b \in M$ und $m \in [0, 1]$

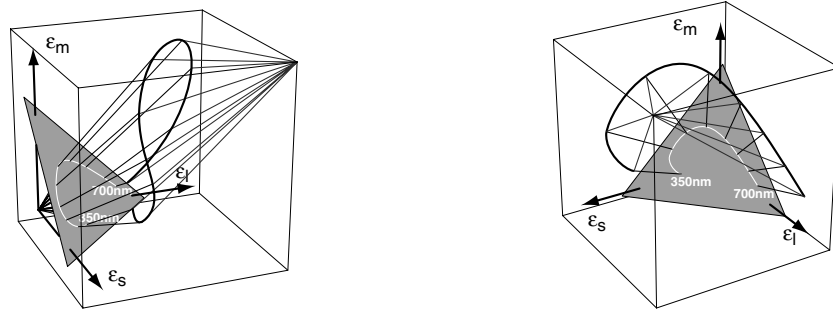


Bild I.4 Rezeptorraum in verschiedenen Ansichten

gen eine Gestalt ähnlich der eines Doppelkegels besitzt. Die Form des isoluminanten Querschnitts entlang konstanter Helligkeit ist nahezu dreieckig. Weiterhin ist ersichtlich, daß im Raum der Rezeptorerregungen kein Koordinatensystem mit orthogonalen Achsen und dem Ursprung $[0, 0, 0]$ existent sein kann.

2.5 Farbverarbeitung im Auge

Da die Absorptionsspektren der Rot- und Grünzapfen sehr ähnlich sind, ist die Aktivität der beiden Zapfentypen stark korreliert. Die Zapfensignale weisen somit neben einer räumlichen auch eine farbliche Kopplung auf. Physiologische Untersuchungen haben ergeben, daß die den Zapfen nachgeschalteten Zellen opponente Kanäle bilden und Farbinformationen in Gegen- und Summenkanälen als Rot-Grün- und Blau-Gelb-Gegensatz übertragen. Die Verrechnung der Farbsignale erfolgt dabei im Netzwerk der Horizontal-, Bipolar-, Amakrin- und Ganglienzellen. Die Verarbeitung bewirkt eine Dekorrelation und Redundanzminderung in den Aktivitäten benachbarter Zapfen.

Durch Summation von Rot und Grün wird ein Gelb-Kanal gebildet, durch Differenz ein Rot-Grün-Kanal. Dies geschieht u.a. durch die hemmende Wirkung der Horizontalzellen auf die zugehörigen Bipolarzellen [Sva56]. Im Gelb-Blau-Kanal schließlich wird die Differenz aus dem Gelb-Kanal und dem Signal der Blauzapfen vorgenommen. Dies erfolgt durch die Amakrinzellen, die eine vergleichbare Funktion wie die Horizontalzellen besitzen (Bild I.5).

Die für die Reizweiterleitung verantwortlichen Ganglienzellen bilden sogenannte parvo- und magnozelluläre Systeme mit koniozellulären Zwischenschichten. Die parvo- und koniozellulären Systeme leiten dabei die Zapfenerregung der Gegenfarbeninformationen Rot-Grün beziehungsweise Gelb-Blau weiter. Zur Erregung des magnozellulären Systems tragen alle drei Zapfentypen additiv bei. Die Systeme unterscheiden sich u.a. hinsichtlich der Übertragungsgeschwindigkeit und der örtlich-zeitlichen Auflösung. Neuronen in der magnozellulären Schicht antworten kurz und vorübergehend auf Reize und sind vergleichsweise schnell. Sie reagieren dabei sehr sensitiv auf Kontraste in ihren rezeptiven Feldern. Die Neuronen sind relativ groß und fassen viele Rezeptoren zusammen, wodurch eine geringere örtliche Auflösung bedingt ist. Durch die Integration vieler Rezeptoren wird jedoch eine hohe zeitliche Auflösung erreicht und das Erkennen von schnellen Helligkeitsschwankungen und Bewegung ermöglicht. Das parvozelluläre System zeigt dagegen langsames und anhaltendes Antwortverhalten und ist selektiv für Veränderungen von Helligkeit und Wellenlängen. Es weist nur geringe Kontrastsensitivität auf. Die Neuronen der für die Blau-Gelb-Differenz zuständigen koniozellulären Schicht antworten ca. 10 ms langsamer als parvozelluläre Neuronen. Durch die kleine Zellgröße der

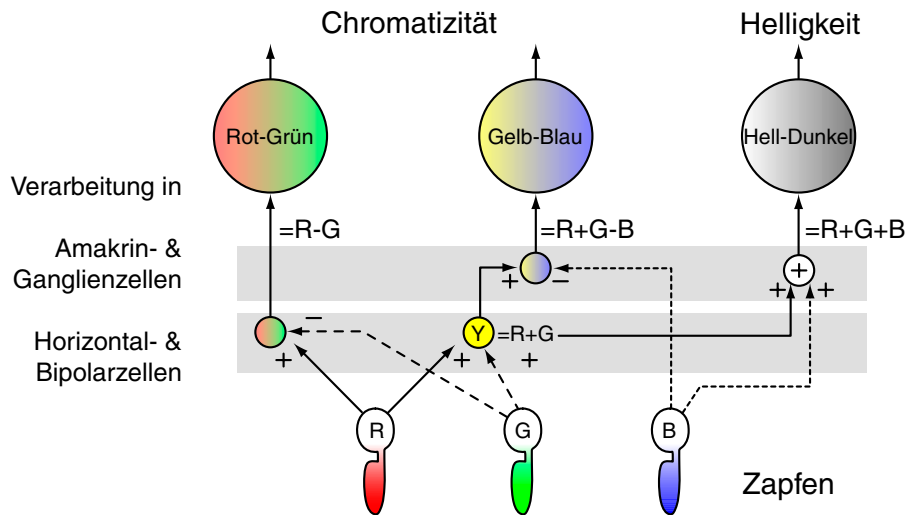


Bild I.5 Farbverarbeitung in der Retina

parvozellulären Neuronen wird eine große räumliche Auflösung erreicht, die jedoch von einer geringeren zeitlichen Auflösung begleitet wird. Parvo- und koniozelluläre Systeme dienen damit hauptsächlich dem Erkennen von Farben, Formen, Texturen und der Tiefenwahrnehmung. Das trichromatische Farbsehen ist ein Beispiel der Populationskodierung, d.h. die Kodierung einer Größe durch die relativen Erregungsstärken unterschiedlich abgestimmter, sich überlappenden Kanäle. Die Separation der Zelltypen setzt sich im LGN (*Lateral Geniculate Nucleus*) und im visuellen Kortex fort. Über die Verarbeitung von Farbe im Gehirn ist bisher jedoch sehr wenig bekannt. Während die Ganglienzellen bevorzugt auf grundsätzliche Farbrichtungen in Gegenfarbkanälen reagieren, finden sich in höheren Hirnarealen andere Farbpräferenzen. So gibt es Hinweise, daß im Gehirn manche Neuronen speziell auf bestimmte Farben ansprechen (bei Makaken) [Drep99].

2.6 Farbadaptation und Umstimmung

Bei andauernden Reizen tritt eine Gewöhnung ein, die bis zum Wegfall der Wahrnehmung führen kann. Wenn die farbempfindlichen Rezeptoren in der Netzhaut längere Zeit dieselbe Farbe wahrnehmen, dann werden sie unempfindlicher und passen sich der Dauerbelastung an. Die entsprechende Farbe wird dann weniger stark wahrgenommen. Beispielsweise kann man auch mit einer getönten Sonnenbrille nach einiger Zeit wieder normale Farben sehen, da sich die Augen an die Farbe der Brille gewöhnen und nicht mehr registrieren. Dieser Vorgang wird Farbadaptation oder Umstimmung genannt und erfolgt vornehmlich bei anhaltenden Reizen. Die Ursache der Gewöhnung kann sowohl eine Verminderung der Empfindlichkeit der Rezeptoren als auch eine veränderte Reaktion des verarbeitenden Nervensystems sein. Entscheidend dabei ist die Reizmenge, d.h. das Produkt aus Reizintensität und Reizdauer [Gold97].

Bei der Wahrnehmung von farbigen Zeichen auf farbigem Grund treten weitere Phänomene auf. Bei bestimmten Farbkombinationen werden an den Objektgrenzen Mischfarben wahrgenommen, z.B. ein weißer Rand bei Blau auf Gelb. Bei kleinen Objekten kann sich die wahrgenommene Farbe in Richtung des Untergrunds verändern (Assimilation), bei großen in Richtung der Komplementärfarbe des Untergrunds (Simultankontrast). Assimilation und Simultankontrast sind auf lokale Adaptationen des Sehsystems zurückzuführen, die das Aussehen von Farben durch den Einfluß der sie umgebenden Farben, den Umfeldfarben, verändern (Bild I.6).

So ist z.B. in der Malerei nicht zu erkennen, wie eine Farbe aussehen wird, wenn sie aus der Tube kommt oder auf der Palette liegt. Das Aussehen ergibt sich erst durch den Einfluß der Umfeldfarben, wenn sie an der entsprechenden Stelle ins Bild gesetzt ist. Das führt u.a. zu der Konsequenz, daß ein Maler zwei verschiedene Farben verwenden muß, damit sie in unterschiedlichem Umfeld gleich aussehen. Andererseits kann es auch sein, daß gleiche Farben verwendet werden müssen, damit sie in verschiedenen Umgebungen verschieden aussehen [TS90].

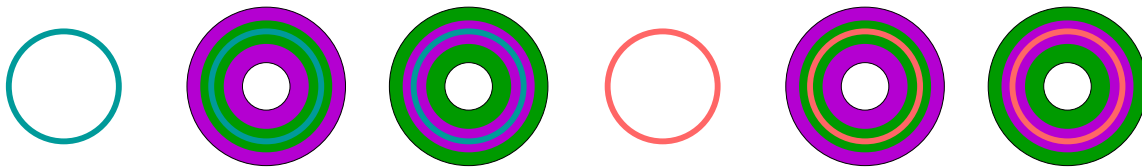


Bild I.6 Beispiel für Assimilation

Um Farben unterscheiden und identifizieren zu können, müssen sie auf achromatischem Hintergrund dargestellt werden. Das gesunde menschliche Auge ist in der Lage, etwa 160 Farben voneinander zu unterscheiden. Werden auch Grautöne, Sättigung und Helligkeiten berücksichtigt, lassen sich etwa 7 Millionen Farbtöne differenzieren [Bös87]. Bei reinen Schwarzweißbildern vermag das Auge nur ca. 30 Grautöne zu unterscheiden [Nie82]. Ohne Vergleichsmöglichkeit können in der Farbbenennung ungeübte Personen jedoch nicht mehr als sechs Farben (Grün, Gelb, Rot, Türkis, Blau und Violett) sowie Unbunt (schwarz-grau-weiß) unterscheiden. Bei Untersuchungen des Wortschatzes unterschiedlicher Sprachen wurde entdeckt, daß im Vokabular zwar unterschiedliche Anzahlen von Farbkategorien vorhanden waren, jedoch nur maximal elf grundlegenden Kategorien existieren, die in allen untersuchten Sprachen zur linguistischen Beschreibung von Farbe dienen. Diese elf grundlegenden Farben sind Rot, Grün, Blau, Gelb, Braun, Violett, Rosa, Orange sowie Schwarz, Weiß und Grau [BK69].

Kapitel II

Modellierung des Farbsehens

朱を奪う紫。

1 Die Modellierung von Licht und Farbe

1.1 Die Definition von Helligkeit

Die Erkenntnisse der Radiometrie, der Strahlungsphysik, ermöglichen die physikalische Erfassung und Meßbarkeit von elektromagnetischer Energie. Sichtbares Licht ist dabei ein schmales Band im elektromagnetischen Spektrum mit Wellenlängen zwischen etwa 400 und 700 nm. Licht, das von der menschlichen Wahrnehmung als „weiß“ wahrgenommen wird, besteht dabei aus einer Mischungen verschiedener Wellenlängen. Die Energie der elektromagnetischen Strahlung kann mit Hilfe eines Spektrometers für jede Wellenlänge gemessen und somit das Spektrum der Strahlung bestimmt werden. In der Photometrie wird die Wirkung des elektromagnetischen Spektrums auf ein visuelles Empfängersystem unter Verwendung physiologischer Aspekte untersucht. Wird beispielsweise eine Lampe von 100 W betrachtet, so sagt diese Größe nicht aus, wie hell diese Lampe auf den Menschen wirkt. Zur Ermittlung der empfundenen Helligkeit wird das Spektrum der Lichtquelle und spektrale Empfindlichkeit der menschlichen Wahrnehmung benötigt. Die spektrale Helligkeitsempfindlichkeit des Menschen wurde 1924 mit dem Internationalen Standard-Beobachter standardisiert (DIN 5031 „Strahlungsphysik“ Teil 2). Dabei existieren unterschiedliche Verläufe der spektralen Hellempfindlichkeit für das photopische Sehen (Tagessehen, $>10 \text{ cd}\cdot\text{m}^{-2}$) $V(\lambda)$, bei dem nur Zapfen angeregt werden, das mesopische Sehen (Dämmerungssehen, $\sim 10 \text{ cd}\cdot\text{m}^{-2}$) $V_{\text{eq}}(\lambda)$ und das skotopische Sehen (Nachtsehen, $<10^{-5} \text{ cd}\cdot\text{m}^{-2}$) $V'(\lambda)$, bei dem nur die farbuntüchtigen Stäbchen angeregt werden (siehe auch Bild I.3 auf Seite 18). Aus dieser spektralen Empfindlichkeitskurve folgt, daß z.B. Licht der Wellenlänge $\lambda=750 \text{ nm}$ (karminrot) bei gleicher Strahlstärke um 10^{-4} weniger hell empfunden wird, als Licht mit der Wellenlänge $\lambda=555 \text{ nm}$ (grün).

1.2 Technische Definition von Farbe

Im Gegensatz zur Helligkeit gestaltet sich die quantitative Erfassung von Farbe jedoch relativ schwierig in Bezug auf objektive Kriterien. Während Helligkeit im wesentlichen durch die Intensität des Lichtes charakterisiert und somit vergleichsweise einfach analysierbar ist, gilt dies für Farben nicht. Farbe ist keine objektiv meßbare physikalische Größe wie Länge oder Gewicht. Zwar kann der Reiz, der zu einer Farbwahrnehmung führt, in Form der elektromagnetischen Strahlung bestimmter Wellenlängen als Licht vermessen werden; die Interpretation, also die eigentliche Farbwahrnehmung, findet erst im Gehirn statt und ist somit einer objektiven Messung nicht zugänglich. Farbe ist folglich ein rein subjektiver Sinneseindruck, für dessen Beschreibung je nach Kontext, z.B. physikalisch, physiologisch, psychologisch oder künstlerisch, völlig unterschiedliche Konzepte existieren. In der DIN Norm 5033 „Farbmessung“ Teil 1 [DIN5033] ist Farbe definiert als: „(...) *diejenige Gesichtsempfindung, durch die sich zwei aneinandergrenzende, strukturlose Teile des Gesichtsfeldes bei einäugiger Beobachtung mit unbewegtem Auge allein unterscheiden können.*“

1.3 Farbmessung und Farbmetrik

Eine Metrik ist allgemein eine Lehre vom Versbau oder Takt und ist mit Maßeinheiten verbunden. Das theoretische Maßkonstrukt ist das Linienelement oder Maßtensor. Die Metrik der Geometrie basiert auf der Maßeinheit, mit der ein Raum zu vermessen ist, dem Meter. Bezogen auf

den Farbraum soll eine Metrik den Unterschied der Farbempfindung mathematisch beschreiben, den zwei beliebige Elemente aus dem Farbraum im Auge des Betrachters unter gegebenen Beobachtungsbedingungen hervorrufen. Das Linienelement des Farbenraumes liefert dabei eine mathematische Beschreibung, mit der Farbunterschiede zu ermitteln sind. Die Idee von Linienelementen geht auf Hermann von Helmholtz zurück, der die wahrnehmbaren Unterschiede zwischen Farbmischungen rechnerisch zu erfassen versuchte. Die Verläufe der Linienelemente hängen jedoch von verschiedenen Messungen, oftmals umstrittenen Annahmen und physikalischen Gesetzen ab. Das Gebiet, das der Physiker Erwin Schrödinger das Terrain der niederen Farbmatrik nannte, baut dabei ausschließlich auf dem Gleichheitsurteil über Farben auf. Die höhere Farbmatrik, die von MacAdam und Stiles begründet wurde, geht dagegen auf die Gleichabständigkeit der Farben ein, die wenig mit deren Gleichheit zu tun hat [FS98]. Das grundlegende Prinzip der Farbmatrik beruht dabei auf der Trichromatizitätstheorie. Diese besagt, daß beim Menschen drei verschiedene retinale Rezeptoren existieren, die Licht über einen begrenzten Wellenlängenbereich absorbieren und empfindlich für die Spektralbereiche Rot, Gelb und Blau sind. In Anhang C, Abschnitt 1 wird eine kurze Zusammenstellung zu der Entwicklung der verschiedenen Farbtheorien gegeben.

2 Die CIE-Farbdefinition

2.1 Vorbetrachtung

Die Vorläufer des CIE⁵-Systems, auf die u.a. in Anhang C, Abschnitt 2 eingegangen wird, weisen Defizite bezüglich Präzision und Objektivität von Farbspezifizierung auf, da Farbmuster benötigt werden. Der Vergleich einer zu bestimmenden Probe oder Lichtquelle mit einem solchen Standard bleibt damit ein subjektiver Vorgang. Des weiteren können die Vergleichsproben im Laufe der Zeit ausbleichen und damit die Zuverlässigkeit der getroffenen Aussagen gefährden. In den ersten Jahrzehnten des 20. Jahrhunderts entstand daher der Wunsch nach einer objektiven Methode zur Farbbestimmung. Daraufhin wurde die CIE beauftragt, eine entsprechende Farbtabelle anzufertigen. Die Ergebnisse liegen seit 1931 vor und haben in ihrer Struktur bis heute Gültigkeit. Das entstandene Farbsystem, auch bekannt unter dem Namen „Normfarbtabelle“ oder „Standard Valenz System“, wurde 1931 zum Standard erklärt [FS98].

Um ein solches Farbsystem entwickeln zu können, bedurfte es verschiedener Grundlagen und Standards. Der wichtigste Standard ist dabei der sogenannte Normalbeobachter. Mit dem Normalbeobachter einigte man sich auf einen physiologischen Standard, der das durchschnittliche Farbwahrnehmungsvermögen des Menschen repräsentiert. Zu jener Zeit war jedoch eine direkte Messung der biologischen Charakteristiken des menschlichen Auges nicht möglich. Der Ausgangspunkt zur Beschreibung der Farbwahrnehmung und zur Ermittlung der hypothetischen Empfindlichkeitskurven $\underline{cie}(\lambda) = [x(\lambda), y(\lambda), z(\lambda)]^T$ des sogenannten Normalbeobachters war daher die indirekte Methode durch Vergleich. Monochromatisches Licht (Spektralfarben) wird vermessen, indem normalsichtige Personen mit einem geeigneten Apparat in die Lage versetzt werden, die Spektralfarbe mit einer additiven Mischung von drei Primärfarben zu vergleichen. Als einzige Forderung besteht dabei, daß die Primärfarben linear unabhängig sind und somit keine der Primärfarben aus den beiden anderen nachgemischt werden kann, wie in Ab-

5. CIE - *Commission Internationale d'Eclairage*

schnitt 2.2.1 gezeigt wird. Im allgemeinen werden dazu Grundfarben mit großem Abstand im Spektrum gewählt, die z.B. Rot, Grün und Blau entsprechen. Bei der Messung durch Vergleich wird dann ermittelt, welche Mischung des gewählten Rot, Grün und Blau der Wahrnehmung von Licht einer bestimmten Wellenlänge entspricht. Die gesuchten Empfindlichkeitskurven ergeben sich dann aus den Mittelwerten der Meßwerte. Die Versuche wurden in den 1920er Jahren von Wright und Gould in England mit einer großen Anzahl menschlicher Beobachter durchgeführt und lieferten die Grundlage für die Standardisierung der Farbmessung. Mit den Farbmischversuchen wurden zwei sogenannte Normspektralwertkurven für ein 2° -Vergleichsfeld und für ein 10° großes Vergleichsfeld festgelegt (Bild II.1). Diese Vergleichsfelder beschreiben dabei den Öffnungswinkel unter dem die Farbreize der Farbprobe auf das Auge des Betrachters auftreffen. Eine Besonderheit des 2° -Beobachters ist, daß bei diesem Betrachtungswinkel nur der Bereich der Fovea gereizt wird. Im Gegensatz dazu sind bei einem 10° -Betrachtungswinkel auch Gebiete außerhalb der Fovea betroffen, wodurch die Farbempfindung besonders im Blaubereich beeinflusst wird. Diese Standards repräsentieren das Farbsehen eines durchschnittlichen normalsichtigen Menschen. Trotz der Umstrittenheit des Normalbeobachters bilden diese Versuchsergebnisse die Basis der modernen Farbmatrik.

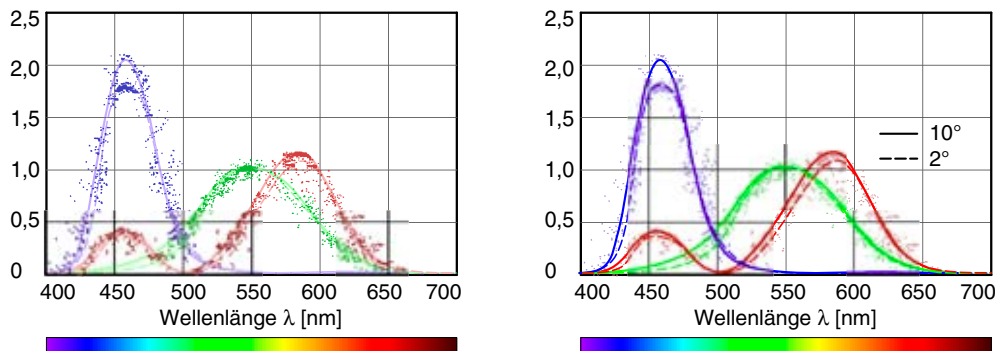


Bild II.1 Meßwerte (links) und Mittelwerte (rechts) der menschlichen Farbempfindlichkeit

Im Bereich der Farbanalyse und -messung werden Formeln üblicherweise einzeln für alle spektralen Empfindlichkeiten, Filterungen etc. angegeben, die hier allgemein als (Farb-) Kanäle bezeichnet werden sollen. Diese separierte Schreibweise ist jedoch für allgemeine Betrachtungen ungünstig, da beispielsweise gemeinsame Rechenoperationen der Kanäle dupliziert und Rechenoperationen zwischen den Kanälen untereinander sehr aufwendig notiert werden müssen. Daher werden im folgenden vektorwertige Funktionen verwendet, die eine einfache Notation der betreffenden Formeln und deren Erweiterung erlauben. An den entsprechenden Stellen ist auf explizite Übergänge zwischen den Schreibweisen hingewiesen.

2.2 Farbmischversuche

Bei einem Farbmischversuch werden die Eigenschaften von Farbmischungen durch indirekte Messung untersucht. Die spektrale Empfindlichkeit $\underline{\vartheta}(\lambda) = [\vartheta_1(\lambda), \vartheta_m(\lambda), \vartheta_s(\lambda)]^T$ der Augen wird dabei als nicht bekannt angenommen. Bei dem Versuch wird üblicherweise ein farbloser Hintergrund mit einem zweiteiligen Feld gezeigt, dessen eine Hälfte eine Farbprobe enthält, während die andere Hälfte von der Mischfarbe gebildet wird. Die Versuchspersonen stellen diese Mischung aus drei gegebenen Primärfarben mit den Spektren $b_1(\lambda)$, $b_2(\lambda)$, $b_3(\lambda)$ mit $\underline{b}(\lambda) = [b_1(\lambda), b_2(\lambda), b_3(\lambda)]^T$ her, indem deren Gewichte $\underline{m} = [m_1, m_2, m_3]^T$

so eingestellt werden, daß kein Unterschied zwischen Probe und Nachmischung wahrgenommen werden kann. Das Spektrum der Mischung ergibt sich dabei zu:

$$\varphi'(\lambda) = m_1 \cdot b_1(\lambda) + m_2 \cdot b_2(\lambda) + m_3 \cdot b_3(\lambda) = \underline{m}^T \cdot \underline{b}(\lambda) \quad (\text{II.1})$$

Wenn kein Unterschied zwischen Probe und Nachmischung wahrnehmbar ist, sind die Erregungsvektoren $\underline{\varepsilon}$ des Auges nach Gleichung (I.2) von Probe und Mischung gleich. Für ein beliebiges Spektrum $\varphi(\lambda)$ der Probe gilt in diesem Fall:

$$\underline{\varepsilon} = \int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot \underline{g}(\lambda) d\lambda = \int_{\lambda=0}^{\infty} \underline{m}^T \cdot \underline{b}(\lambda) \cdot \underline{g}(\lambda) \cdot d\lambda \quad (\text{II.2})$$

Der Vektor \underline{m} wird dabei als Dreifarbwert (*tristimulus value*) des von den gewählten Primärvalenzen $\underline{b}(\lambda)$ gebildeten Farbraumes bezeichnet. Formal sind auch negative Komponenten des Dreifarbwertes möglich. Im Experiment bedeutet dies, daß das betreffende Licht nicht dem Abgleichfeld, sondern der Probe zugemischt wird. Die Komponente geht dabei mit positivem Vorzeichen in den Term der Erregungen ein. Dies ist jedoch der Berücksichtigung mit negativem Vorzeichen auf der rechten Seite äquivalent. Ist die Probe eine normierte reine Spektralfarbe der Wellenlänge λ_0 , so vereinfacht sich die Erregung der Zapfen zu:

$$\underline{\varepsilon} = \int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot \underline{g}(\lambda) d\lambda = \int_{\lambda=0}^{\infty} \delta(\lambda - \lambda_0) \cdot \underline{g}(\lambda) \cdot d\lambda = \underline{g}(\lambda_0) \quad (\text{II.3})$$

Auch in diesem Fall sind die Rezeptorerregungen für die reine Spektralfarbe und die Nachmischung gleich und es gilt mit Ersetzung von $\underline{m} = \underline{a}(\lambda_0)$ in Gleichung (II.2):

$$\underline{\varepsilon} = \underline{g}(\lambda_0) = \int_{\lambda=0}^{\infty} \underline{a}(\lambda_0)^T \cdot \underline{b}(\lambda) \cdot \underline{g}(\lambda) \cdot d\lambda \quad (\text{II.4})$$

Die von den Versuchspersonen eingestellten Dreifarbwerte $\underline{a}(\lambda)$, die sich nun als Funktion der Wellenlänge ergeben, werden Spektralwertfunktion genannt. Durch Einsetzen von $\underline{g}(\lambda_0)$ nach Gleichung (II.4) in die Erregungsfunktion des Auges nach Gleichung (II.2) erhält man:

$$\underline{\varepsilon} = \int_{\lambda_0=0}^{\infty} \varphi(\lambda_0) \cdot \underline{g}(\lambda_0) \cdot d\lambda_0 = \int_{\lambda_0=0}^{\infty} \varphi(\lambda_0) \cdot \left[\int_{\lambda=0}^{\infty} \underline{a}(\lambda_0)^T \cdot \underline{b}(\lambda) \cdot \underline{g}(\lambda) \cdot d\lambda \right] \cdot d\lambda_0 \quad (\text{II.5})$$

Durch Vertauschen der Integration folgt:

$$\underline{\varepsilon} = \int_{\lambda=0}^{\infty} \left[\int_{\lambda_0=0}^{\infty} \varphi(\lambda_0) \cdot \underline{a}(\lambda_0) \cdot d\lambda_0 \right]^T \cdot \underline{b}(\lambda) \cdot \underline{g}(\lambda) \cdot d\lambda \quad (\text{II.6})$$

Aus dem Vergleich mit Gleichung (II.2) ergibt sich nun:

$$\underline{m} = \int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot \underline{a}(\lambda) \cdot d\lambda \quad (\text{II.7})$$

Der Dreifarbwert \underline{m} eines beliebigen Spektrums $\varphi(\lambda)$ entspricht somit dem Integral über dessen Produkt mit der Spektralwertfunktion $\underline{a}(\lambda)$.

2.2.1 Mathematische Betrachtung der Farbwahrnehmung

Zur Vereinheitlichung der verschiedenen existierenden Modellierungen von Farbaufnahme und -wiedergabe wurde ein neues Beschreibungssystem entwickelt, das im folgenden als Grundlage der daraus abgeleiteten Anwendungen dient.

Bei der Aufnahme von Farben und deren Wiedergabe handelt es sich formal gesehen um eine Integraltransformation mit entsprechender Rücktransformation. Das Spektrum des aufgenommenen Lichtes $\varphi(\lambda)$ kann als eine kontinuierliche Funktion der Wellenlänge λ aufgefaßt wer-

den, die durch eine Spektralwertfunktion genannte Transformationsfunktion $a(\lambda, \kappa)$, über die Beziehung $m(k) = \int_{\lambda=-\infty}^{\infty} \varphi(\lambda) \cdot a(\lambda, \kappa) \cdot d\lambda$ in eine kontinuierliche Funktion $m(\kappa)$ transformiert wird. Mit der inversen Transformation kann das originale Spektrum idealerweise zu $\varphi(\lambda) = \int_{\kappa=-\infty}^{\infty} m(k) \cdot b(\lambda, \kappa) \cdot d\kappa$ zurückgewonnen werden. Die explizite Bestimmung der zur Rücktransformation notwendigen Funktion $b(\lambda, \kappa)$ ist dabei ein grundlegendes Problem der Theorie der Integraltransformationen und soll hier nicht allgemein, sondern nur für den Spezialfall der Farbwahrnehmung betrachtet werden.

Im Fall der Farbaufnahme wird keine Transformation mit kontinuierlichem Basisfunktionsraum verwendet. Üblicherweise werden nur wenige Funktionen $b_k(\lambda)$ mit $k \subset \kappa$ verwendet, in der Regel beschränkt man sich auf $|k| = 3$. λ ist zudem physikalisch nur positiv definiert. Die Koeffizienten der Transformation ergeben sich dann zu $m_k = \int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot a_k(\lambda) \cdot d\lambda$ oder in Vektorschreibweise zu $\underline{m} = \int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot \underline{a}(\lambda) \cdot d\lambda$. Unter diesen Bedingungen vereinfacht sich das Integral der Rücktransformation zu $\varphi'(\lambda) = \sum_k m_k \cdot b_k(\lambda)$ mit der Vektorschreibweise $\varphi'(\lambda) = \underline{m}^T \cdot \underline{b}(\lambda)$.

Zur Ermittlung von geeigneten Funktionen $b_k(\lambda)$ zur Rücktransformation gibt es prinzipiell zwei mögliche Wege. Als erstes soll folgende Annahme verwendet werden: Wird das Ergebnis der Rücktransformation $\varphi'(\lambda)$ erneut in die Hintransformation eingesetzt, so sollen sich identische Koeffizienten \underline{m} ergeben:

$$\underline{m} = \int_{\lambda=0}^{\infty} (\underline{m}^T \cdot \underline{b}(\lambda)) \cdot \underline{a}(\lambda) \cdot d\lambda = \left[\int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \underline{b}(\lambda)^T \cdot d\lambda \right] \cdot \underline{m} \quad (\text{II.8})$$

In Bezug auf die Darstellung von Farben bedeutet dies, daß die Aufnahme der Nachmischung einer Farbe zu demselben Dreifarbwert führt, mit dem die Nachmischung erzeugt wurde. Dies stellt die Beschreibung einer konsistenten Wiedergabe dar und ergibt die Bedingung:

$$\int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \underline{b}(\lambda)^T \cdot d\lambda = \underline{E} \quad (\text{II.9})$$

Mathematisch kann dies als Orthogonalitätsbedingung interpretiert werden und stellt einen fundamentalen Zusammenhang dar. Die Koordinate i von $\underline{a}(\lambda)$ steht senkrecht auf den Komponenten j von $\underline{b}(\lambda)$ mit $i \neq j$. Auf die Mischung von Farben bezogen bedeutet dies, daß die Nachmischung einer Primärfarbe $b_k(\lambda)$ nur aus der Grundfarbe selbst bestehen kann. Die Forderung nach der linearen Unabhängigkeit der verwendeten Primärfarben für Farbmischversuche in Abschnitt 2.1 leitet sich direkt hieraus ab.

Als zweiter Weg zur Ermittlung von Rücktransformationfunktionen $b_k(\lambda)$ wird das Ergebnis der Hintransformation zurücktransformiert: $\varphi'(\lambda) = \sum_k \left(\int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot a_k(\lambda) \cdot d\lambda \right) \cdot b_k(\lambda)$. In Vektorschreibweise ergibt sich damit vereinfacht:

$$\varphi'(\lambda) = \left(\int_{\lambda=0}^{\infty} \varphi(\lambda) \cdot \underline{a}(\lambda) \cdot d\lambda \right)^T \cdot \underline{b}(\lambda) \quad (\text{II.10})$$

Dabei soll das resultierende Spektrum $\varphi'(\lambda)$ mit dem originalen Spektrum $\varphi(\lambda)$ übereinstimmen. Zur Auflösung des Integrals wird angenommen, daß diese Spektren mit einer Basisfunktion $\underline{c}(\lambda)$ zu $\varphi'(\lambda) = \underline{\Phi}'^T \cdot \underline{c}(\lambda)$ und $\varphi(\lambda) = \underline{\Phi}^T \cdot \underline{c}(\lambda)$ vollständig beschrieben sind:

$$\begin{aligned} \underline{\Phi}'^T \cdot \underline{c}(\lambda) &= \left(\int_{\lambda=0}^{\infty} \underline{\Phi}^T \cdot \underline{c}(\lambda) \cdot \underline{a}(\lambda) \cdot d\lambda \right)^T \cdot \underline{b}(\lambda) \\ \underline{\Phi}'^T \cdot \underline{c}(\lambda) &= \underline{\Phi}^T \cdot \left(\int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \underline{c}(\lambda)^T \cdot d\lambda \right)^T \cdot \underline{b}(\lambda) \end{aligned} \quad (\text{II.11})$$

Wird Identität von $\underline{\Phi}'$ und $\underline{\Phi}$ verlangt, ergibt sich die Bedingung für die Transformation zu:

$$\begin{aligned}\underline{c}(\lambda) &= \left(\int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \underline{c}(\lambda)^T \cdot d\lambda \right)^T \cdot \underline{b}(\lambda) \\ \underline{b}(\lambda) &= \left(\int_{\lambda=0}^{\infty} \underline{c}(\lambda) \cdot \underline{a}(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{c}(\lambda)\end{aligned}\quad (\text{II.12})$$

Für den Fall, daß die Basisfunktion zu $\underline{c}(\lambda) = \underline{a}(\lambda)$ gewählt werden, ist die Rücktransfunktionsfunktion $\underline{b}(\lambda)$ folglich bestimmt mit:

$$\underline{b}(\lambda) = \left(\int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \underline{a}(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{a}(\lambda) \quad (\text{II.13})$$

die gleichzeitig auch die Bedingung nach Gleichung (II.9) erfüllt. Wird die Basisfunktion hingegen mit $\underline{c}(\lambda) = \underline{b}(\lambda)$ festgelegt, ergibt sich aus Gleichung (II.12):

$$\underline{a}(\lambda) = \left(\int_{\lambda=0}^{\infty} \underline{b}(\lambda) \cdot \underline{b}(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{b}(\lambda) \quad (\text{II.14})$$

Die zwei Gleichungen zur Ermittlung der jeweils inversen Transformationsfunktion sind gleichwertig, jedoch wegen der Festlegung auf ein spezifisches Basissystem nicht gleichzeitig gültig. Im Bezug auf die Farbverarbeitung kann mit Gleichung (II.13) aus gegebenen Spektralwertfunktionen eines Sensors der entsprechende Primärfarbraum und mit Gleichung (II.14) aus einer Primärfarbbeschreibung die Empfindlichkeit des äquivalenten Sensors bestimmt werden. Da für die Rücktransformation jedoch mit $\phi'(\lambda) = \underline{\Phi}'^T \cdot \underline{c}(\lambda)$ nur Linearkombinationen der Basisfunktion $\underline{c}(\lambda)$ zur Verfügung stehen, ist bei endlicher Dimensionalität des Basisraumes nur für Spezialfälle eine ideale Rekonstruktion des Eingangsspektrums $\phi(\lambda)$ zu erreichen.

Natürlicher Betrachtungsvorgang

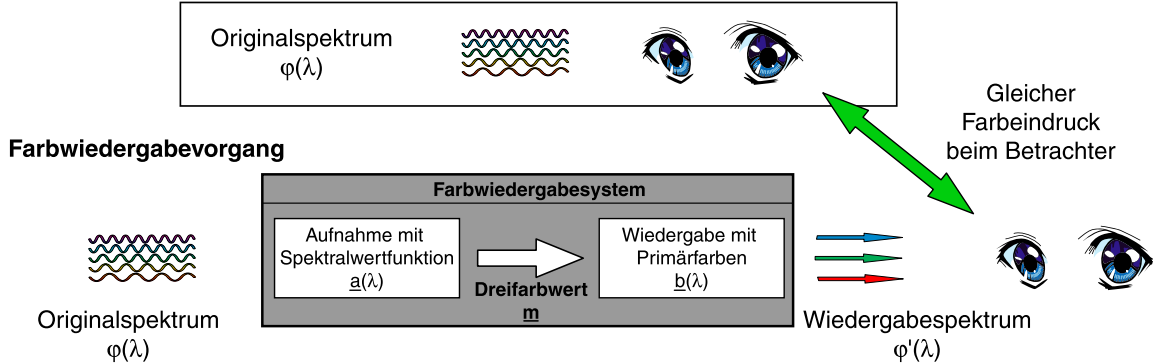


Bild II.2 Prinzip eines Farbwiedergabesystems

Für ein Farbwiedergabesystem nach Bild II.2 ist die exakte Rekonstruktion des Eingangsspektrums $\phi(\lambda)$ zudem nicht notwendig. Aufgrund der Metamerie ist es ausreichend, ein Wiedergabespektrum $\phi'(\lambda) = \underline{m}^T \cdot \underline{b}(\lambda)$ zu erzeugen, das im Auge den gleichen Farbeindruck hervorruft wie das originale Spektrum. Praktische Primärfarben sind dabei beliebig wählbar, sofern die Komponenten linear unabhängig sind. Der für die Nachmischung mit den Primärfarben $\underline{b}(\lambda)$ benötigte Dreifarbwert \underline{m} und die zugehörige Spektralwertfunktion $\underline{a}(\lambda)$ sind in diesem Fall jedoch durch einen Farbmischversuch nach Gleichung (II.4) bzw. Gleichung (II.7) direkt an die menschliche Farbwahrnehmung gebunden. In Kapitel III, Abschnitt 1 soll diese Beziehung als Grundlage der Farbmessung und Farbkorrektur im Detail ausgeführt werden.

Die Primärfarbfunktion $\underline{b}(\lambda)$ und die äquivalente Spektralwertfunktion $\underline{a}(\lambda)$ stellen somit vektorwertige Funktionen dar, die sich zur Darstellung von Farbwerten eignen. Ein Dreifar-

wert \underline{m} beschreibt die Koordinaten einer bestimmten Farbe in diesem von Primärfarbfunktion und Spektralwertfunktion gebildeten Farbsystem B . Je nach Wahl der Primärfarben ergeben sich dabei unterschiedliche Spektralwertfunktionen und andere Werte für die Farbkoordinaten. Die Primärfarben stellen eine konstruktive Basis dar, mit der ein Spektrum $\varphi'(\lambda) = \underline{m}^T \cdot \underline{b}(\lambda)$ einer zu beschreibenden Farbe gebildet wird. Die Spektralwertfunktionen entsprechen einer deskriptiven Basis, mit deren Hilfe die Farbwerte \underline{m} eines beliebigen Spektrums $\varphi(\lambda)$ mit $\underline{m} = \int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$ zu finden sind. Primärfarben und Spektralwertfunktion verhalten sich zueinander wie kovariante und kontravariante Koordinaten in der affinen Geometrie.

Formal existieren unendlich viele Farbwiedergabesysteme, die jedoch äquivalent sind: Gegeben seien zwei verschiedene Farbwiedergabesysteme B_A und B_B mit den Primärfarben $\underline{b}_A(\lambda)$ und $\underline{b}_B(\lambda)$ sowie den zugehörigen Spektralwertfunktionen $\underline{a}_A(\lambda)$ und $\underline{a}_B(\lambda)$. Bei der Reproduktion einer Farbkoordinate \underline{m}_A in B_A wird das zugehörige Spektrum $\varphi'_A(\lambda) = \underline{m}_A^T \cdot \underline{b}_A(\lambda)$ gebildet, das beim Menschen einen spezifischen Farbeindruck hervorruft. Um mit dem Farbsystem B_B einen identischen Farbeindruck zu erzeugen, wird eine Farbkoordinate \underline{m}_B benötigt, die sich aus der Projektion von $\varphi'_A(\lambda)$ auf die Spektralwertfunktion $\underline{a}_B(\lambda)$ ergibt:

$$\underline{m}_B = \int_{\lambda=0}^{\infty} \underline{a}_B(\lambda) \cdot \varphi'_A(\lambda) \cdot d\lambda = \int_{\lambda=0}^{\infty} \underline{a}_B(\lambda) \cdot \underline{m}_A^T \cdot \underline{b}_A(\lambda) \cdot d\lambda = \int_{\lambda=0}^{\infty} \underline{a}_B(\lambda) \cdot \underline{b}_A(\lambda)^T \cdot d\lambda \cdot \underline{m}_A \quad (\text{II.15})$$

Die Farbwerte verschiedener Primärfarbsysteme sind somit durch eine lineare Transformation verbunden. Die Transformationsmatrix $\underline{\mathfrak{Z}}_{AB} = \int_{\lambda=0}^{\infty} \underline{a}_B(\lambda) \cdot \underline{b}_A(\lambda)^T \cdot d\lambda$ ergibt sich dabei direkt aus den Parametern der Farbsysteme. Die Transformation ist umkehrbar, es gilt $\underline{m}_B = \underline{\mathfrak{Z}}_{AB} \cdot \underline{m}_A$ und $\underline{m}_A = \underline{\mathfrak{Z}}_{AB}^{-1} \cdot \underline{m}_B$.

Zur Umrechnung verschiedener Primärfarbsysteme ohne Kenntnis deren Spektralwertfunktionen ist die Transformationen in einen Farbraum B_X mit bekannter Spektralwertfunktion $\underline{a}_X(\lambda)$ möglich: $\underline{\mathfrak{Z}}_{AX} = \int_{\lambda=0}^{\infty} \underline{a}_X(\lambda) \cdot \underline{b}_A(\lambda)^T \cdot d\lambda$, $\underline{\mathfrak{Z}}_{BX} = \int_{\lambda=0}^{\infty} \underline{a}_X(\lambda) \cdot \underline{b}_B(\lambda)^T \cdot d\lambda$. Die Transformation zwischen \underline{m}_A und \underline{m}_B ergibt sich dann zu: $\underline{m}_B = \underline{\mathfrak{Z}}_{BX}^{-1} \cdot \underline{\mathfrak{Z}}_{AX} \cdot \underline{m}_A$ bzw. $\underline{m}_A = \underline{\mathfrak{Z}}_{AX}^{-1} \cdot \underline{\mathfrak{Z}}_{BX} \cdot \underline{m}_B$. Da diese Transformationen jedoch im Allgemeinen nicht orthonormal sind, bleiben beim Wechsel zwischen verschiedenen Primärfarbsystemen Abstände und Winkel nicht erhalten. Somit ist im Farbenraum eine feste Definition von Abständen und Winkeln nicht sinnvoll, da kein Primärfarbsystem ausgezeichnet ist. Zusätzlich ergibt sich aus der Orthogonalitätsbedingung zwischen Primärfarben und Spektralwertfunktionen nach Gleichung (II.9), daß diese nicht gleichzeitig im Raum der physikalisch möglichen Spektren, der nur aus positiven Werten besteht, auftreten können. Wird von realisierbaren Primärfarben ausgegangen, so weisen die zugehörigen Spektralwertfunktionen stets negative Werte auf. Transformiert man dagegen das Koordinatensystem zu positiven Spektralwertfunktionen, so existieren keine realen Primärfarben für eine direkte Nachmischung.

2.3 Die CIE Spektralwertfunktion

Das Ziel des CIE-Systems war eine möglichst objektive Beschreibung der Funktionalität des menschlichen Auges. Daher wurden für das CIE-System zwei realisierbare Spektralwertfunktionen als Basis gewählt, die sogenannten Normalbeobachter $\underline{cie}_{2^\circ}(\lambda)$ und $\underline{cie}_{10^\circ}(\lambda)$. Eine vollständige Beschreibung des Auges gelang mit dieser Festlegung zwar nicht, jedoch zeigen Wyszecki und Stiles [WS82], daß ein linearer Zusammenhang zwischen den gewählten Normalbeobachtern und den spektralen Absorptionsfunktionen der Farbrezeptoren des Auges be-

steht. Der Nachweis, daß alle realisierbaren Farbräume untereinander und auch mit dem Raum der biologischen Rezeptorerregungen äquivalent sind, wird in Kapitel III, Abschnitt 1 im Rahmen der Farbkorrekturmethode erbracht.

Die spektralen Sensitivitäten $\underline{cie}(\lambda) = [x(\lambda), y(\lambda), z(\lambda)]^T$ sind über ihrem Definitionsbereich nicht-negativ und können z.B. durch Farbfilter realisiert werden. Die zugehörigen Primärfarben sind jedoch unreal, da sie für bestimmte Wellenlängen negative Abstrahlungen erfordern (Bild II.3). Es kann somit kein Anzeigegerät existieren, das die Ausgaben der Normalbeobachter direkt verwendet. Da jedoch die Transformation zwischen den verschiedenen Primärfarbsystemen möglich ist, kann zur Anzeige ein beliebiger Farbraum mit realisierbaren Primärfarben gewählt werden. Die Matrix zur Transformation ergibt sich dabei aus den spektralen Eigenschaften des Anzeigesystems nach Abschnitt 2.2.1. Realisierbare Primärfarbsysteme verwenden in der Praxis Rot, Grün und Blau als Basis und werden daher abkürzend RGB-Systeme genannt. Aufgrund der vielfältigen Möglichkeiten zur Erzeugung von praktischen Primärfarben existieren viele verschiedene RGB-Systeme, die jedoch nur für ein bestimmtes Anzeigegerät verwendbar sind. Ein allgemeingültiges, ausgezeichnetes RGB-System existiert nicht.

Mit den Kurven der Normalbeobachter lassen sich bei der Farbaufnahme der Normfarbvektor $\underline{m}_{cie} = [X, Y, Z]^T$ und die Normfarbwerte X , Y und Z einer spezifischen spektralen Leistungsdichte $\varphi(\lambda)$ mit $[X, Y, Z]^T = \int_{\lambda=0}^{\infty} [x(\lambda), y(\lambda), z(\lambda)]^T \cdot \varphi(\lambda) \cdot d\lambda$ bestimmen. Die XYZ-Tristimuluswerte können dabei alle in der Natur möglichen Farben mit positiven Koeffizienten beschreiben. Die spektralen Gewichtungsfunktionen des Normalbeobachters sind standardisiert [DIN5033] und ermöglichen, Farben in einem einheitlichen Koordinatensystem festzulegen sowie Farbinformationen ohne Farbmuster allein durch Zahlen zu übermitteln. Aus den Normfarbwerten werden alle weiteren farbmtrischen Kennwerte und Farbräume, wie beispielsweise CIE $L^*a^*b^*$ und CIE $L^*u^*v^*$ abgeleitet (siehe auch Anhang C, Abschnitt 2). Theoretisch könnten die zur Ermittlung der Spektralwertfunktion verwendeten Primärfarben beliebige Spektralreize sein, sofern sie linear unabhängig sind. Um jedoch stets die gleiche Bezugsbasis zu schaffen, wurden die monochromatischen Primärvalenzen von der CIE ebenfalls einheitlich festgelegt. Dies sind: Rot: $X = 700,0 \text{ nm}$, relative Strahldichte⁶ = 72,0962, Grün: $Y = 546,1 \text{ nm}$, relative Strahldichte = 1,3791 und Blau: $Z = 435,8 \text{ nm}$, relative Strahldichte = 1,0000. Zugleich legte die CIE fest, daß zu vermessende Farbproben unter der durchschnittlichen Tagesbeleuchtung, später als CIE Standardleuchtkörper D65 definiert, zu betrachten sind. Andere definierte Normlichtarten sind beispielsweise CIE Illuminant A für Glühlampenlicht, D50 und D55 für Tageslicht bei Durchsicht- respektive Aufsichtvorlagen, die in Abschnitt 3.4 näher beschrieben werden. Ein bestimmtes Blau besitzt somit bezogen auf die Normlichtart D65 und ein 2°-Gesichtsfeld einen spezifischen Normfarbwert XYZ. Die Standards entstanden jedoch auf Basis von empirischen Auswertungen, denen nicht zuletzt wegen der zwingend auftretenden Messungenauigkeiten eine gewisse Willkür nachgesagt wird. Trotzdem haben sie sich etabliert und bilden heute die Basis für die Farbmessung. Eine vollständige Definition des CIE Systems ist in der Publikation [CIE86] veröffentlicht.

6. Die Strahldichte ist als Strahlungsleistung eines Flächenelementes in Richtung eines infinitesimalen Raumwinkels definiert (siehe auch Anhang A, Abschnitt 1.2.8).

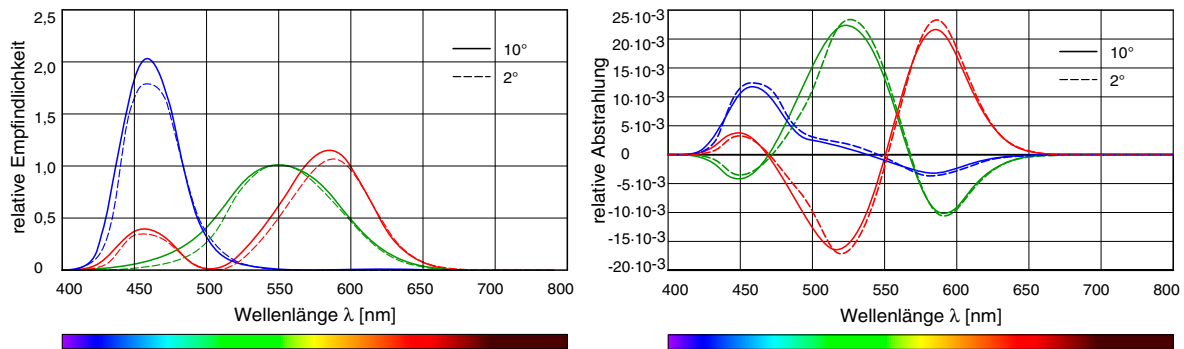


Bild II.3 CIE-Normalbeobachter (links) und zugehörigen Primärfarben (rechts)

2.4 Das CIE-System

Das CIE-System baut auf dem Normalbeobachter, den Normspektralwertkurven und dem resultierenden Dreifarbwert XYZ auf. Ein oft geäußelter Nachteil des XYZ-Farbraumes ist, daß aus den Werten X, Y und Z nicht direkt ableitbar ist, wie der Farbreiz in Wirklichkeit aussieht. Dies ist insofern gerechtfertigt, da Farbton und Sättigung nicht ohne weiteres aus den Farbwerten abgelesen werden können. Der Verlauf von $y(\lambda)$ wurde jedoch entsprechend dem spektralen Helligkeitsempfinden des menschlichen Auges gewählt (Abschnitt 1.1, Kapitel I, Abschnitt 2.1). Der Normfarbwert Y entspricht somit der wahrgenommenen Helligkeit, während die Komponenten X und Z das Farbempfinden repräsentieren. Der von X, Y und Z definierte dreidimensionale Raum kann um eine Dimension reduziert werden, um eine zweidimensionale Farbkarte zu gewinnen. Dies geschieht durch Projektion des Farbraumes auf eine Ebene (Bild II.4). Dazu werden die gemessenen Farbwerte X, Y und Z in Farbwertanteile x , y und z umgerechnet:

$$x = \frac{X}{(X + Y + Z)}, \quad y = \frac{Y}{(X + Y + Z)}, \quad z = \frac{Z}{(X + Y + Z)} \quad (\text{II.16})$$

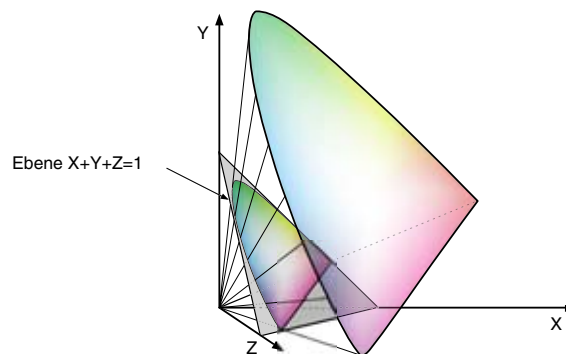


Bild II.4 Projektion des CIE-XYZ-Raumes auf die Ebene $X+Y+Z=1$

Da $x + y + z = 1$ gilt, ist die Angabe der Koordinate z nicht notwendig, da sie bereits durch x und y definiert ist. In einer Ebene eingetragen sind die unabhängigen Größen (x, y) zur Definition einer beliebigen Farbe ausreichend (Bild II.5). Zur vollständigen Beschreibung bedarf es jedoch noch einer dritten Komponente. Allgemein wird dazu der Normfarbwert Y verwendet, mit dem die Helligkeit der Farbprobe definiert ist. Im (x, y) -Diagramm ist die wahrgenommene Helligkeit in x -Richtung konstant. Der sich so ergebende Raum ist der CIE-xyY-Farbraum. Die Umkehrung der Rechnung ergibt sich nach [Poy96] zu:

$$X = \frac{x}{y} \cdot Y, \quad Z = \frac{1 - x - y}{y} \cdot Y \quad (\text{II.17})$$

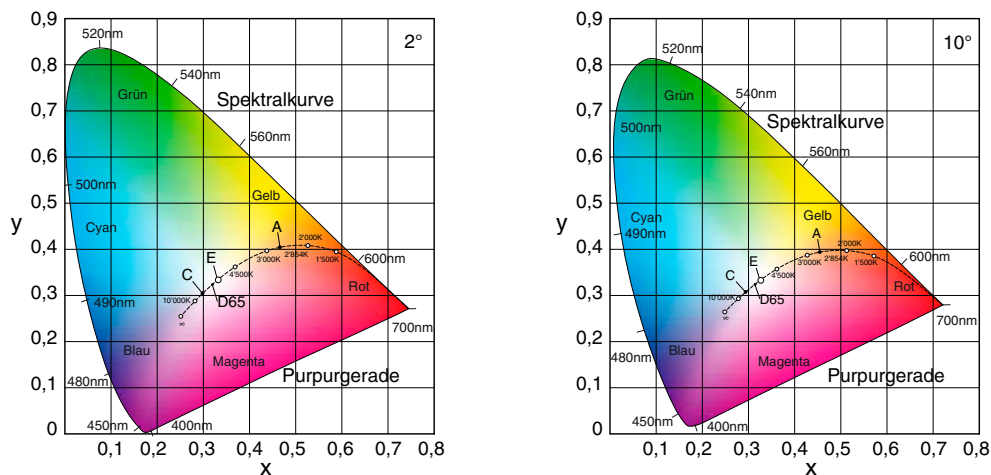


Bild II.5 CIE-xyY-Diagramme für den 2° und 10° Beobachter mit Farbtemperaturbogen (siehe auch Abschnitt 3.4)

Die Form des CIE-Standarddiagrammes wird somit von den Wertepaaren (x, y) für die Farben bestimmt, die aus nur einer Wellenlänge bestehen, den monochromatischen bzw. Spektralfarben. Dieser Spektralfarbenzug bildet eine nicht geschlossene Kurve. Die Verbindungsgerade zwischen Anfang (400 nm) und Ende (700 nm) des Spektralfarbenzuges ist die Purpurgerade. Auf ihr befinden sich die Mischfarben zwischen Anfang und Ende des sichtbaren Spektrums. Innerhalb der Spektralkurve und der Purpurgerade liegen sämtliche Farbvalenzen, die sich durch additive Mischung von Spektralfarben erzielen lassen. Diese bilden die Gesamtheit aller realen, für das menschliche Auge wahrnehmbaren Farben. Farben außerhalb dieses Bereichs sind aufgrund der sich überlappenden Komponenten der Spektralwertfunktion nicht realisierbar, da kein zugehöriges physikalisches Spektrum existiert.

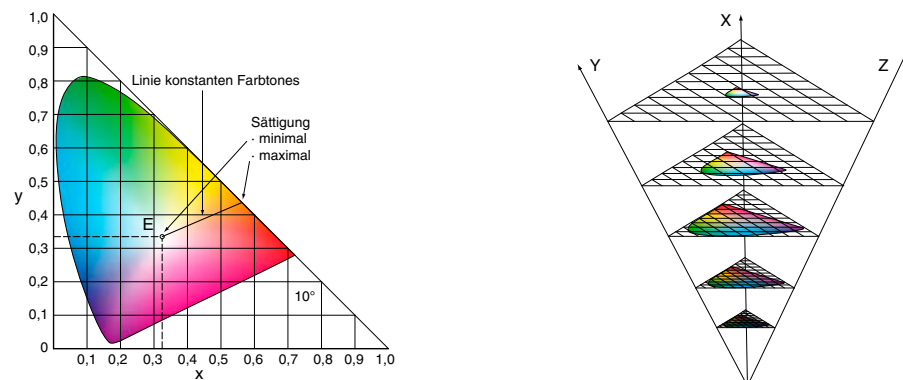


Bild II.6 CIE-xy-Diagramm und Schnitte für verschiedene Helligkeiten

Wie bei der Purpurgeraden liegen die realen Farben, die aus je zwei Spektralfarben gemischt werden, auf einer Linie zwischen den beiden entsprechenden Punkten des Spektralfarbenzugs. Der genaue Wert ist dabei von der Gewichtung der Intensitäten der beiden Wellenlängen abhängig. Die Gesamtheit aller Linien bildet die konvexe Hülle des Spektralfarbenzugs. Der Spektralfarbenzug selbst befindet sich innerhalb des rechtwinkligen und gleichschenkligen Dreiecks mit den Eckpunkten $(x, y) : (0,0), (0,1), (1,0)$. Der Mittelpunkt des Dreiecks liegt bei den Normfarbwertanteilen $(x, y) : (0.333, 0.333)$. Diese Farbvalenz heißt daher Mittelpunktvalenz und wird mit dem Buchstaben E bezeichnet. Der Punkt mit der geringsten Farbintensität, der sogenannten Sättigung, wird als Unbuntpunkt (Weiß-Punkt) bezeichnet. Im CIE-Normvalenzsystem sind Mittelpunktvalenz E und Unbuntpunkt identisch. Nach außen nimmt im Dia-

gramm die Sättigung zu. Die Farbtöne mit der höchsten Sättigung sind in der CIE-Farbtabelle auf dem Spektralfarbenzug zu finden. Dieser stellt die reinen Farben dar. Wird ein beliebiger Punkt auf dem Spektralfarbenzug durch eine Gerade mit der Mittelpunktsvalenz verbunden, so sind auf dieser Gerade sämtliche Mischfarben zwischen der Spektralfarbe und der Unbuntvalenz (Weiß) zu finden. Der Farbton dieser Mischfarben ist identisch, jedoch nimmt die Sättigung in Richtung des Unbuntpunktes ab. Somit kann der Abstand einer Farbvalenz zum Unbuntpunkt als ein Maß für die Sättigung betrachtet werden [Schäp93] (Bild II.6).

Das CIE-Diagramm stellt jedoch nur eine Ebene des Farbraumes dar. Die anderen Ebenen repräsentieren Farben abnehmender Helligkeit und zeigen, wie die Farben aussehen, wenn weniger Licht vorhanden ist. Zur Berücksichtigung der Helligkeit in der Farbdefinition wird die zur Fläche senkrechte Koordinate Y verwendet. Für die wenigsten Farbtöne des CIE-Diagramms kann dabei der Maximalwert von $Y = 100$ erreicht werden. So kann die Helligkeit des Spektralfarbenzuges höchstens 4.68 betragen. Je ungesättigter jedoch eine Farbe ist, desto höher kann die Helligkeit sein; für die Mittelpunktsvalenz erreicht sie definitionsgemäß den Wert 100. Mit dem CIE-System ist die weltweite Verständigung über Farben möglich geworden. Der Empfänger einer Farbprobe kann mit Hilfe der drei Zahlenwerte x , y , Y den jeweiligen Farbort lokalisieren, ohne die zugehörige Farbe gesehen zu haben. Seit Bestehen des Systems hat sich auch seine uneingeschränkte Eignung für alle technischen Anwendungen der Farbe bis hin zum Farbfernsehen erwiesen.

2.5 Nachteile des CIE-Systems

Das größte Defizit des CIE-Systems ist die Verletzung des Kriteriums der Gleichabständigkeit bezüglich der Wahrnehmung. Ein System ist gleichabständig bezüglich der Wahrnehmung, wenn die Störung eines Wertes über dem gesamten Wertebereich gleichmäßig wahrnehmbar ist. Beispielsweise muß ein Lautstärkesteller, der gleichabständig bezüglich der Wahrnehmung ist, so entworfen sein, daß eine Drehung um einen festen Winkel stets einem identischen Wahrnehmungssprung der Lautstärke über dem gesamten Einstellbereich entspricht. Da die menschliche Wahrnehmung von Lautstärke logarithmische Natur besitzt, muß der Lautstärkesteller ebenfalls logarithmisch sein. Eine lineare Steuerung würde die Wahrnehmung zwischen Stellwinkel und empfundener Lautstärke verzerren.

Auch im Bereich der optischen Wahrnehmung ist das lineare System CIE-XYZ weit von einer Gleichabständigkeit bezüglich der Wahrnehmung entfernt. Die Koordinaten X , Y , Z erlauben die Entscheidung, ob zwei Farben unter gegebenen Randbedingungen übereinstimmen. Aus den CIE-Normfarbwertanteilen x und y , sowie dem Normfarbwert Y lassen sich dazu die sogenannten Helmholtz-Maßzahlen (DIN 5033 Teil 3) ableiten, anhand derer beurteilt werden kann, ob zwei gegebene Farben gleich oder verschieden sind. Falls jedoch keine Übereinstimmung besteht, ist nicht ersichtlich, wie groß der wahrnehmbare Unterschied zwischen zwei Farben ist. Die Differenz von zwei XYZ-Werten ist ebenfalls kein geeignetes Maß. Somit ist es nicht möglich, empfindungsgleiche Farbabstände, d.h. Farbdifferenzen aus dem CIE-XYZ-System abzulesen. Physiologisch gleichabständige Farben sind nicht gleichabständig dargestellt.

Wird beispielsweise ein bestimmtes Blau und ein Grün als Farbmuster gewählt und jeweils ein zweites Muster, die nach subjektivem Farbempfinden in beiden Fällen gleich weit vom gewählten Grün oder Blau entfernt sind, stellt man beim Eintragen der zugehörigen xy -Werte in das

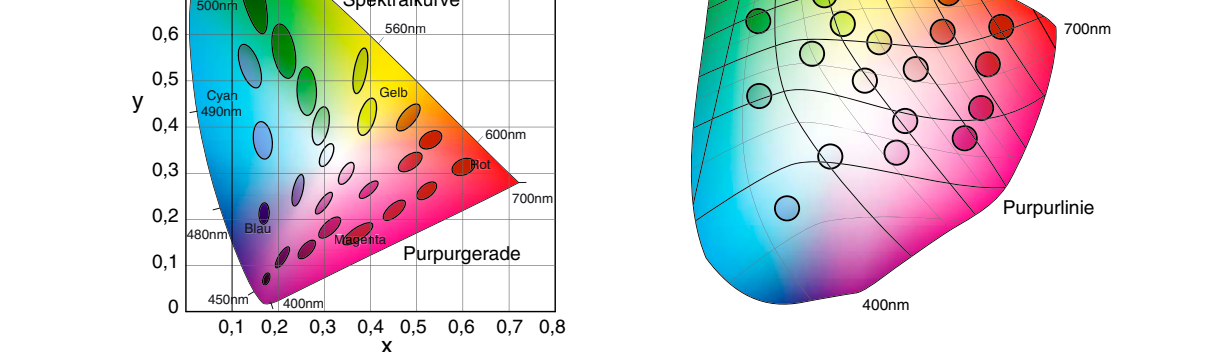
CIE-Diagramm fest, daß die beiden grünen Muster im Diagramm deutlich weiter auseinander liegen als die blauen Muster. Die Erklärung ist im Wahrnehmungsvorgang zu finden. Blaues Licht leistet zwar einen verhältnismäßig geringen Beitrag zur Helligkeitsempfindung, jedoch besitzt die menschliche Wahrnehmung eine außerordentlich gute Fähigkeit zur Farbunterscheidung im blauen Bereich. So können bei blauen Farben bereits geringe Variationen wahrgenommen werden, während im roten und besonders im grünen Bereich auch größere Unterschiede unterhalb der Wahrnehmungsschwelle bleiben. Die Diskrepanz zwischen dargestellter und wahrgenommener Farbabständigkeit spiegelt sich auch in der uneinheitlichen Raumaufteilung des CIE-Diagramms bezüglich der verschiedenen Farbkategorien wieder, die sich darin äußert, daß Farben mit Grünanteil (Gelb, Grün, Cyan) überrepräsentiert und die roten, violetten und blauen Töne zusammengestaucht sind.

2.5.1 MacAdam-Schwellenellipsen

Der Farbabstand ΔE von zwei Farbproben soll den empfundenen Unterschied zwischen Farbvalenzen durch eine Maßzahl beschreiben. Maßgebliche Erkenntnisse zum Farbabstand lieferten die Studien von MacAdam [MacA42]. Der Psychophysiker Douglas L. MacAdam hatte sich als erster systematisch mit der Farbabständigkeit beschäftigt und damit nicht nur die Gleichheit, sondern auch die Gleichabständigkeit von Farbe betrachtet und dadurch die Entwicklungen der Farbabstandsformeln und somit die Meßbarkeit von Farbabständen entscheidend vorangetrieben. Allgemeine Untersuchungen zur Farbunterschiedsempfindlichkeit waren bis dahin undenkbar, denn schon allein die Tatsache, daß jede Farbe zu allen ihren Nachbarn im Farbenraum verschieden große Schwellenwerte besitzt, ließ es aussichtslos erscheinen, in diesem Bereich zu relevanten und brauchbaren Ergebnissen zu kommen.

Für seine Versuche wählte MacAdam 25 verschieden Farborte innerhalb der Normfarbtafel aus. Die Farbarten wurden mittels Filter erzeugt und besaßen alle die gleiche Leuchtdichte von ca. $50 \text{ cd} \cdot \text{m}^{-2}$. Die Farbvalenzen wurde der Versuchsperson auf einer Gesichtshälfte dargeboten. Auf der anderen Gesichtshälfte sollte die Versuchsperson die vorgegebene Farbvalenz aus zwei anderen Farbvalenzen nachmischen, die so gewählt waren, daß ihre Verbindungsgerade in der Farbtafel durch den Farbort der vorgegebenen Farbvalenz ging und diese exakt nachmischen konnte. Die Nachmischung wurde von der Versuchsperson mehrfach wiederholt, wobei die für die Mischung verwendeten Primärfarben unterschiedlich waren. Die ermittelten Farborte ordneten sich aufgrund der Einstellfehler ellipsenförmig um die vorgegebenen Farbvalenz an (Bild II.7). Diese experimentell und empirisch ermittelten Ellipsen zeigen somit die Farborte an, innerhalb der von der Sollfarbe nicht zu unterscheidenden Farben liegen. Die Ellipsen stellen somit eine Schwelle dar, ab der eine zur Sollfarbe abweichende Farbe erkannt werden kann.

Dabei sind sowohl Größe als auch Lage der Ellipsen innerhalb der Normfarbtafel unterschiedlich. Dies verdeutlicht, daß die Darstellung des CIE-XYZ-Farbraumes empfindungsgemäß stark verzerrt ist. Wäre das System gleichabständig, müßten gleichabständige Farborte gleich weit von der definierten Sollfarbe entfernt auf einem Kreis liegen. Ebenfalls müßten die Kreise unabhängig von der gewählten Sollfarbe gleich groß sein. Jeweils gleichabständige Farborte der gegebenen Sollfarben liegen jedoch nicht auf Kreisen, sondern auf Ellipsen. Zusätzlich liegen die gleichabständigen Farborte zu einer gegebenen Sollfarbe je nach Sollfarbe auf verschiedenen großen Ellipsen. Spätere Versuche, die auch Farbunterschiede durch Variation der Leuchtdichte



Erst Ende der 50er Jahre wurde die Festlegung einer Farbabstandsformel in das Arbeitsprogramm der CIE-Kommission „Farbmessung“ aufgenommen. Da man keine neuen experimen-

3 Reproduktion von Farbe

Die Reproduktion von Farbeindrücken ist aufgrund der Metamerie relativ einfach möglich. Bereits mit drei unabhängigen Basisfarben lassen sich alle möglichen Farbtöne nachbilden. Für die überwiegende Anzahl von Anwendungen ist dies ausreichend. Die Nachbildung kompletter Spektren zur Reproduktion von Farben ist nur in wenigen Spezialfällen notwendig und wird beispielsweise bei der Rekonstruktion historischer Gemälde verwendet [TB88]. Die metamere Nachbildung eines Farbeindruckes kann durch zwei verschiedene Mechanismen geschehen, der additiven und der subtraktiven Farbmischung (Bild II.8).

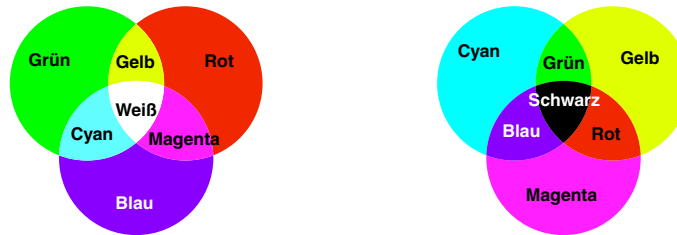


Bild II.8 Additive (links) und subtraktive Farbmischung (rechts)

3.1 Additive Farbmischung

Die additive Farbmischung beruht auf der spektralen Überlagerung unterschiedlicher Lichtquellen. Diese liegt beispielsweise vor, wenn Bilder aus mehreren Projektoren mit je verschiedenfarbigem Licht überblendet werden. Das Spektrum des Ergebnisses besteht dabei aus der Summe der Einzelspektren. Sind $b_1(\lambda)$ und $b_2(\lambda)$ die spektralen Leistungsdichten zweier Einzelfarben mit den zugehörigen Farbvektoren \underline{m}_1 und \underline{m}_2 , so ergibt sich der Dreifarbwert der additiven Mischung bei Gewichtung mit g_1 und g_2 zu:

$$\begin{aligned} \underline{m}_{12} &= \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot (g_1 b_1(\lambda) + g_2 b_2(\lambda)) \cdot d\lambda \\ &= g_1 \cdot \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot b_1(\lambda) \cdot d\lambda + g_2 \cdot \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot b_2(\lambda) \cdot d\lambda \\ &= g_1 \cdot \underline{m}_1 + g_2 \cdot \underline{m}_2 \end{aligned} \quad (\text{II.18})$$

Dies bedeutet anschaulich, daß die Valenz der additiven Mischfarbe nur von den Valenzen und den Gewichten, nicht jedoch von der spektralen Zusammensetzung der beteiligten Farben abhängig ist (Gesetz von Grassmann).

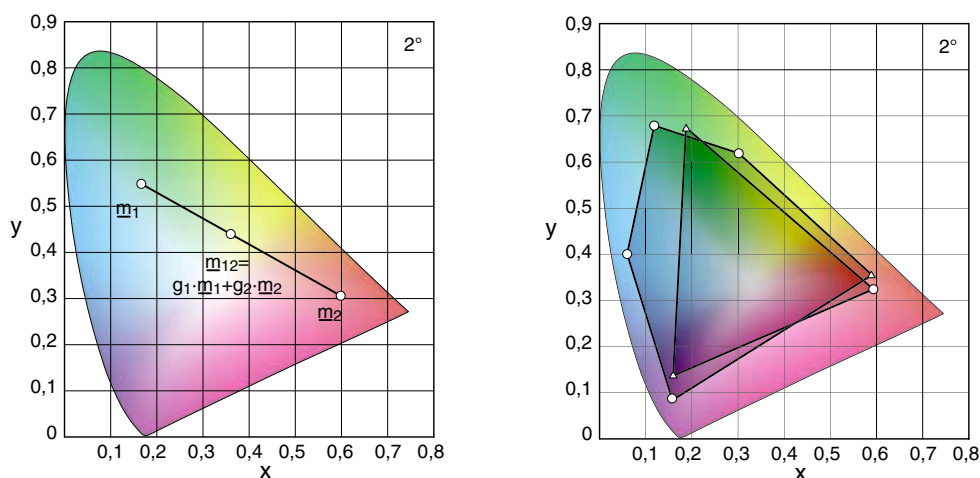


Bild II.9 Additive Farbmischung im CIE-Diagramm
Beispiele für die Mischung von zwei Farbprimären (links) und den mischbarer Farbberreich von drei sowie fünf Farbprimären (rechts)

Führt man in Gleichung (II.18) $s = g_1 + g_2$ und $t = g_1 / (g_1 + g_2)$ ein, so ergibt sich:

$$\underline{m}_{12} = s \cdot (t \cdot \underline{m}_1 + (1 - t) \cdot \underline{m}_2) = s \cdot ((\underline{m}_1 - \underline{m}_2) \cdot t + \underline{m}_2) \quad (\text{II.19})$$

Unter der Voraussetzung, daß die Farbdarstellung mit dem Normfarbvektor zu $\underline{m} = [X, Y, Z]^T$ erfolgt, führt die Berechnung der Farbwertanteile x_{12} und y_{12} nach Gleichung (II.16) zu folgender Darstellung:

$$\begin{aligned} x_{12} &= \frac{(X_1 - X_2) \cdot t + X_2}{((X_1 + Y_1 + Z_1) - (X_2 + Y_2 + Z_2)) \cdot t + (X_2 + Y_2 + Z_2)} \\ y_{12} &= \frac{(Y_1 - Y_2) \cdot t + Y_2}{((X_1 + Y_1 + Z_1) - (X_2 + Y_2 + Z_2)) \cdot t + (X_2 + Y_2 + Z_2)} \\ y_{12} &= \underbrace{\frac{(X_1 + Y_1 + Z_1) \cdot Y_2 - (X_2 + Y_2 + Z_2) \cdot Y_1}{(X_1 + Y_1 + Z_1) \cdot X_2 - (X_2 + Y_2 + Z_2) \cdot X_1}}_{\alpha} \cdot x_{12} + \underbrace{\frac{Y_1 \cdot X_2 - Y_2 \cdot X_1}{(X_1 + Y_1 + Z_1) \cdot X_2 - (X_2 + Y_2 + Z_2) \cdot X_1}}_{\beta} \end{aligned} \quad (\text{II.20})$$

Diese Geradengleichung entspricht in der CIE-Farbtabelle der Verbindungslinie zwischen den Farbarten der Farben \underline{m}_1 und \underline{m}_2 . Diese Erkenntnis kann zur Beschreibung des Verhaltens von Reproduktionsmedien benutzt werden, die auf der additiven Farbmischung basieren. Das CIE-Diagramm eignet sich hierfür, da in dieser Darstellung die Linearität beibehalten wird. Da diese Darstellung jedoch keine Gleichabständigkeit von Farben aufweist, existiert keine lineare Beziehung zwischen den Farbarten auf dieser Linie (Bild II.9).

Eine Mischung von drei linear unabhängigen Farben öffnet eine dreieckige Fläche innerhalb des Spektralfarbzuges, die alle wahrnehmbaren Farben enthält, die sich durch unterschiedliche Gewichtung dieser drei einzelnen Primärvalenzen ergeben. Die Mischfarben eines Systems mit vier oder mehr Grundfarben liegen in einem durch diese Farben aufgespannten Polygon, das der konvexen Hülle der Primärfarben entspricht. Zur Ermittlung des Farbeindrucks einer Mischung lassen sich jeweils zwei beliebige Primärvalenzen schrittweise zusammenfassen, bis sich der resultierende Farbeindruck als einzelner Punkt im Farbraum ergibt.

Von praktischer Bedeutung ist die additive Farbmischung durch optische Verschmelzung kleiner Bildpunkte, wie sie z.B. bei Farbbildschirmen, beim Offsetdruck, beim Weben oder in der pointillistischen Malerei verwendet wird. Werden diese Objekte aus genügend großem Abstand betrachtet, so addieren sich die Spektren der örtlich getrennt erzeugten Quellen auf der Retina des Betrachters. Eine andere Anwendung ist der Farbkreis, bei dem farbige Flächen durch die Drehung in schnellem Wechsel gezeigt werden, so daß eine temporale Mischung erfolgt. Bereits Isaac Newton demonstrierte, daß die Oberfläche einer schnell rotierenden Scheibe, die sektorweise blau, grün und rot gefärbt war, grau erschien, sobald die Rotationsgeschwindigkeit groß genug war.

3.2 Subtraktive Farbmischung

Der additiven Farbmischung steht die subtraktive Farbmischung gegenüber. Subtraktive Farbmischung tritt u.a. auf, wenn der Farbeindruck durch Absorption bestimmt wird. Pigmente oder Farbstoffe verändern dabei die spektrale Zusammensetzung eines eintreffenden Lichtes, indem sie unabhängig voneinander Teile des Lichtes absorbieren. Dies wird durch ein Absorptions-

spektrum beschrieben. Die sich ergebende Farbe wird im Fall von Körperfarben als spektraler Reflexionsgrad bezeichnet [Schöp93]. Beispiele sind das Vermischen von Pigmenten für die Malerei oder beim Farbdruck.

Ist der spektrale Absorptionsgrad einer Probe definiert mit $\alpha(\lambda)$, so ist bei Körperfarben der spektrale Reflexionsgrad oder Reflektanz $\rho(\lambda)$ bestimmt durch: $\rho(\lambda) = 1 - g \cdot \alpha(\lambda)$, mit dem Gewicht $g \in [0, 1]$, welches die Deckkraft der Farbe auf einem ideal reflektierendem Untergrund bestimmt. Bei durchscheinenden farbigen Materialien kann die Transmission $\tau(\lambda)$ analog zur Reflektanz betrachtet werden. Das von der Probe reflektierte bzw. transmittierte Licht ist dabei komplementär zum absorbierten Licht. Das zur Wahrnehmung führende Lichtspektrum $\varphi(\lambda)$ wird jedoch mit $\varphi(\lambda) = u(\lambda) \cdot \rho(\lambda) = u(\lambda) \cdot (1 - g \cdot \alpha(\lambda))$ zusätzlich vom Umgebungslicht $u(\lambda)$ mit bestimmt. Der zur Beschreibung des Farbeindrucks verwendete Farbwert ergibt sich aus der Integration des Lichtspektrums mit dem Normalbeobachter:

$$\begin{aligned} \underline{m} &= \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot \varphi(\lambda) \cdot d\lambda = \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot u(\lambda) \cdot (1 - g \cdot \alpha(\lambda)) \cdot d\lambda \\ &= \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot u(\lambda) \cdot d\lambda - g \cdot \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot u(\lambda) \cdot \alpha(\lambda) \cdot d\lambda \\ &= \underline{m}_u - g \cdot \underline{m}_{u\alpha} \end{aligned} \quad (\text{II.21})$$

Im CIE-Diagramm liegen somit der Farbort des Umgebungslichtes \underline{m}_u , der Farbort der beleuchteten Probe $\underline{m}_{u\alpha}$ und die sich ergebende Farbe \underline{m} auf einer Geraden.

Bei der subtraktiven Farbmischung zweier Farbsubstanzen ist der resultierende Farbeindruck im Gegensatz zur additiven Farbmischung von der spektralen Zusammensetzung der einzelnen Komponenten abhängig. Werden zwei Farben subtraktiv gemischt, so multiplizieren sich deren Reflektanzen, da beide Farbanteile unabhängig voneinander Teile des einstrahlenden Lichtspektrums absorbieren. Das Spektrum des wahrgenommen Lichtes ergibt sich somit zu:

$$\begin{aligned} \varphi_{12}(\lambda) &= u(\lambda) \cdot \rho_1(\lambda) \cdot \rho_2(\lambda) = u(\lambda) \cdot (1 - g_1 \cdot \alpha_1(\lambda)) \cdot (1 - g_2 \cdot \alpha_2(\lambda)) \\ &= u(\lambda) \cdot (1 - g_1 \cdot \alpha_1(\lambda) - g_2 \cdot \alpha_2(\lambda) + g_1 \cdot g_2 \cdot \alpha_1(\lambda) \cdot \alpha_2(\lambda)) \end{aligned} \quad (\text{II.22})$$

Der resultierende Farbwert berechnet sich demzufolge mit:

$$\begin{aligned} \underline{m}_{12} &= \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot \varphi_{12}(\lambda) \cdot d\lambda \\ &= \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot u(\lambda) \cdot (1 - g_1 \cdot \alpha_1(\lambda) - g_2 \cdot \alpha_2(\lambda) + g_1 \cdot g_2 \cdot \alpha_1(\lambda) \cdot \alpha_2(\lambda)) \cdot d\lambda \\ &= \int_{\lambda} \underline{cie}(\lambda) \cdot u(\lambda) \cdot d\lambda - g_1 \cdot \int_{\lambda} \underline{cie}(\lambda) \cdot u(\lambda) \cdot \alpha_1(\lambda) \cdot d\lambda - g_2 \cdot \int_{\lambda} \underline{cie}(\lambda) \cdot u(\lambda) \cdot \alpha_2(\lambda) \cdot d\lambda \\ &\quad + g_1 \cdot g_2 \cdot \int_{\lambda} \underline{cie}(\lambda) \cdot u(\lambda) \cdot \alpha_1(\lambda) \cdot \alpha_2(\lambda) \cdot d\lambda \end{aligned} \quad (\text{II.23})$$

Wird als weitere Annahme das Umgebungslicht mit gleichförmigen Leistungsdichtespektrum definiert, so vereinfacht sich Gleichung (II.23) zu:

$$\underline{m}'_{12} = \underline{1} - g_1 \cdot \underline{m}_1 - g_2 \cdot \underline{m}_2 + g_1 \cdot g_2 \cdot \int_{\lambda=0}^{\infty} \underline{cie}(\lambda) \cdot \alpha_1(\lambda) \cdot \alpha_2(\lambda) \cdot d\lambda \quad (\text{II.24})$$

Eine Mischfarbe kann bei subtraktiver Farbmischung somit nur dann unabhängig von den spektralen Eigenschaften der Einzelfarben bestimmt werden, wenn sich die Absorptionsspektren der Einzelfarben $\alpha_1(\lambda)$ und $\alpha_2(\lambda)$ nicht überlappen und $\alpha_1(\lambda) \cdot \alpha_2(\lambda) = 0$ gilt. Die möglichen Mischfarben \underline{m}_{12} bilden im vereinfachten Fall ein Polygon im CIE-Diagramm, dessen Eck-

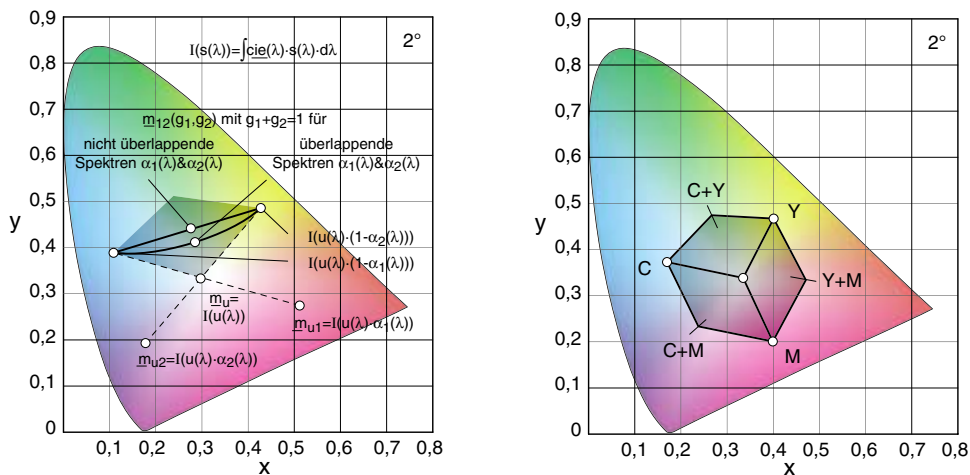


Bild II.10 Prinzip der subtraktiven Farbmischung im CIE-Diagramm
 mischbarer Farbbereich von zwei (links) und von drei Farbprimären (rechts)

punkte aus dem Unbuntpunkt E und den Farben $\underline{1} - \underline{m}_1$, $\underline{1} - \underline{m}_2$, sowie $\underline{1} - \underline{m}_1 - \underline{m}_2$ bestehen (Bild II.10). Praktische Druckerfarben und Tinten besitzen allerdings spektrale Absorptionskurven, die sich erheblich überdecken. Die Druckfarbe Magenta absorbiert beispielsweise neben den vorgesehen mittleren Wellenlängen (grün) auch einen Teil des kurzwelligen (blauen) Lichtes. Bei der Reproduktion einer Farbe, welche die Absorption des gesamten kurzwelligen Lichtes erfordert, z.B. Braun, hat diese beiläufige Absorption des Magenta keinen Einfluß auf die Farbwiedergabe. Für andere Farben enthält die Reproduktion jedoch zu wenig Blau und erscheint somit besonders bei mittleren Tönen gelbstichig. Ähnliche Interaktionen existieren auch zwischen den anderen Primären praktischer Farben und Tinten. Formal ergibt sich damit kein linearer Zusammenhang zwischen Einzelfarben und Mischfarbe bei praktischen Implementierungen der subtraktiven Farbmischung. Die Farborte der Mischfarben liegen in diesen Fällen im CIE-Diagramm auf gekrümmten Kurven. Dennoch wird in der Praxis häufig ein linearer Zusammenhang angenommen. Diese Methode hat sich besonders bei einfachen Anwendungen durchgesetzt, die keine exakte Reproduktion von Farben benötigen. Die damit zwangsläufig auftretenden Fehler, die sich als Falschfarben äußern, werden dabei mehr oder weniger ignoriert. Der Druck und die photographische Wiedergabe werden zusätzlich durch die Nichtlinearitäten der drei (oder mehr) Farbprimäre erschwert. Beispielsweise verursachen beim Offsetdruck die physikalischen Prozesse eine als Tonwertzunahme bezeichnete Nichtlinearität. Ein typisches Drucksystem erzeugt beispielsweise bei Druck mit halber Aussteuerung eine graue Fläche mit einem Reflexionsvermögen von ca. 0.26 und nicht, wie erwartet von 0.5. Eine Berechnung der Druckfarben kann demnach nicht sinnvoll durchgeführt werden, ohne diese Nichtlinearitäten in Betracht zu ziehen. Die Transformation von kolorimetrischen Bilddaten für Printmedien ist jedoch in der Literatur hinreichend beschrieben [SCB88].

3.3 Anwendung der additiven Farbmischung

Die aktuell bedeutendsten Reproduktionsmittel für Farben und Farbbilder sind Bildschirme und Monitore. Diese beruhen auf der additiven Farbmischung von üblicherweise drei Primärfarben in Rot, Grün und Blau und sind mit äquivalenten RGB-Farbsystemen beschrieben. Zur Erzeugung der Primärfarben sind verschiedene Techniken im Einsatz. In Kathodenstrahlröhren werden Elektronen auf phosphoreszierende Schichten gelenkt und damit zum Leuchten angeregt.

Bei LCDs⁷ werden Farbfilter verwendet, die aus dem Licht einer zusätzlichen Lichtquelle oder aus dem Umgebungslicht die entsprechenden Grundfarben ausfiltern. Das eigentliche LCD vermag nur die Transmission durch den Flüssigkristall zu steuern, indem das Licht entsprechend der Ansteuerung polarisiert und mit einem Polfilter proportional zur Polarisierung ausgefiltert wird. Bei der DLP⁸ oder auch DMD⁹ Technologie geschieht diese Steuerung der Strahlstärke des Lichtes durch die Reflexion an digital angesteuerten Spiegeln. Aktuell sind auch Bildschirme auf Basis von Leuchtdioden in Entwicklung, wobei Realisierungen mit organischen LEDs aufgrund der günstigen Herstellung am aussichtsreichsten sind.

Die Gemeinsamkeit aller dieser Realisierungen liegt in der additiven Farbmischung. Wenn dabei drei Grundfarben verwendet werden, ergibt sich der vom Bildschirm wiedergebbare Farbraum im CIE-Farbdigramm als ein Dreieck. Die Koordinaten der Eckpunkte werden dabei von den Grundfarben bestimmt. Da die Primärfarben in der Praxis im allgemeinen roten, grünen und blauen Bereich liegen, spricht man auch von RGB-Systemen. Zur Maximierung eines solchen Wiedergabefarbraumes sind sehr schmalbandige Primärfarben erforderlich, deren Koordinaten im CIE-Farbdigramm nahe der Spektralkurve liegen. Eine weitere Vergrößerung kann durch zusätzliche Farbprimäre erreicht werden. Dafür stehen prinzipiell jedoch nur zwei mögliche Bereiche zur Verfügung: An der Verbindungsachse von Rot und Grün kann der Farbraum durch eine gelbe Primärfarbe und an der Blau-Grün-Achse durch Cyan zu einem Fünfeck erweitert werden. Durch diese zusätzlichen Primärfarben lassen sich besonders Farben im gelben und blaugrünen Bereich wesentlich besser darstellen. Für die Verbindungslinie von Blau und Rot existiert keine zugehörige Spektralfarbe, daher kann an dieser Stelle keine Erweiterung des Wiedergaberaumes erfolgen. In [Gui04] ist ein Display vorgestellt, das mit fünf Primärfarben eine subjektiv intensivere Farbdarstellung ermöglicht als Wiedergabegeräte, die nur mit drei Primärfarben arbeiten. Yamaguchi et al. präsentieren in [YTO02] ein System mit sechs Farbprimären, die durch Aufspaltung der üblicherweise für die Wiedergabe mit drei Farben verwendeten Farbprimäre erhalten werden (Bild II.11).

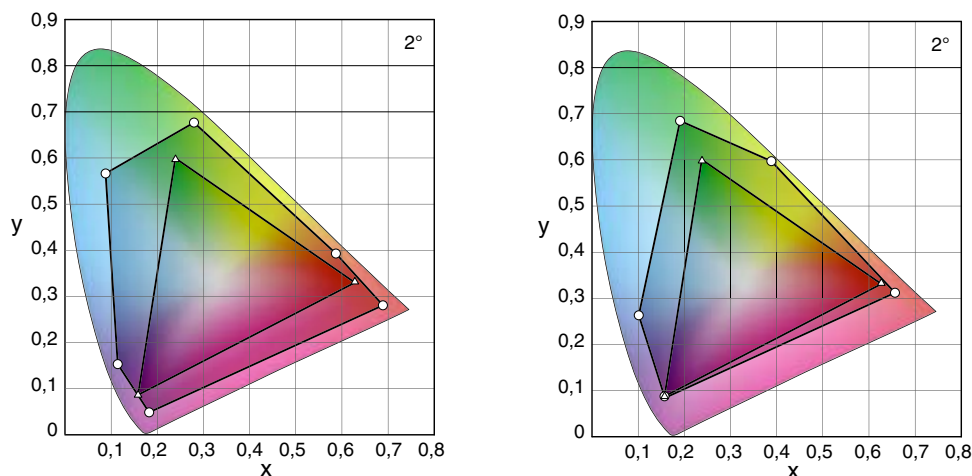


Bild II.11 Farbumfang des Sechs-Farb- und Fünf-Farb-Displays im Vergleich zu einem herkömmlichen Dreifarbbildschirm, approximierte Orte der Primärfarben nach [YTO02] (links) und Abbildung nach [Gui04] (rechts)

7. LCD - *Liquid Crystal Device* - Flüssigkristall-(Gerät)

8. DLP - *Digital Light Processing* - Digitale Lichtverarbeitung

9. DMD - *Digital Micromirror Device* - Digital-Spiegel-(Gerät)

Aus diesen neuen Entwicklungen von Wiedergabemedien erwachsen auch Herausforderungen für Farbbildsensoren. Die neuen Sensoren müssen in der Lage sein, Farben über dem gesamten, vom Menschen wahrnehmbaren Farbbereich aufzunehmen und zu verarbeiten. Üblicherweise werden die Einschränkungen des wiedergebbaren Farbraumes als Grund genannt, die Farbaufnahme nur auf bestimmte Bereiche des wahrnehmbaren Farbraumes zu beschränken [Poy95].

3.4 Weißpunkt

Bei der additiven Bildwiedergabe mit einem RGB-System ist der Weißpunkt als diejenige Farbe definiert, die sich durch gleiche Anteile an roten, grünen und blauen Bestandteilen ergibt. Der Weißpunkt ist demzufolge eine Funktion des Verhältnisses der Energien der Farbprimäre. Bei der subtraktiven Wiedergabe wird der Weißpunkt bestimmt durch das Leistungsdichtespektrum des Umgebungslichtes, multipliziert mit der Reflektanz des unbedruckten Mediums. Es gibt somit keine physikalische oder wahrnehmungstechnische Definition von Weiß. Der Weißpunkt wird bei der Wiedergabe explizit spezifiziert. Für eine einfache Berechnung bietet sich die Annahme eines gleichförmigen Leistungsdichtespektrum an, die als „equal-energy illuminant“ oder CIE Illuminant E bekannt ist. Eine realistischere Annahme zur Approximation von Tageslicht ist als Leuchtkörper D65 spezifiziert worden. Die Definition dieser Leuchtkörper kommt aus der physikalischen Betrachtung sogenannter Schwarzkörperstrahler. Max Planck (1858-1947) stellte fest, daß das Leistungsdichtespektrum eines heißen Gegenstandes eine Funktion von dessen Temperatur ist (Bild II.12). Da technische Lichtquellen in der Regel einen heißen Kern besitzen, ist es sinnvoll, die Leuchtkörper zu klassifizieren, indem die Temperatur in Kelvin [K] eines Schwarzkörperstrahlers mit identischer Farbe angegeben wird. Ein Leuchtkörper kann somit durch Angabe der äquivalenten Farbtemperatur formlos spezifiziert werden.

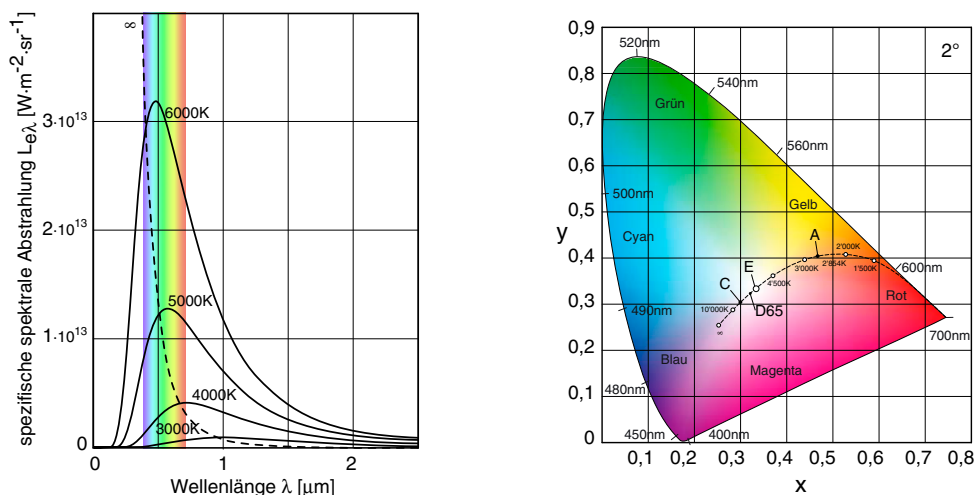


Bild II.12 Spektralverläufe von verschiedenen Planckschen Schwarzkörperstrahlern (links) und deren Farborte im CIE-Diagramm (rechts)

CIE Standard Illuminant A stellt den Leuchtkörper dar, der einem Schwarzkörperstrahler von 2854 K entspricht. Die Realisierung dieses Leuchtkörpers ist eine gasgefüllte Wolframfadlampe mit äquivalenter Arbeitstemperatur. Die CIE-Standardleuchtkörper B und C waren ursprünglich als Modell für direktes Sonnenlicht angelegt und besitzen äquivalente Farbtemperaturen von 4874 K bzw. 6774 K. Aufgrund von schwerwiegenden Unzulänglichkeiten besonders im UV-Bereich, der u.a. wichtig für fluoreszierende Materialien ist, werden die

Modelle nicht mehr verwendet. Die Leuchtkörper der D-Reihe werden mit der entsprechenden Farbtemperatur angegeben. Der Leuchtkörper D65 entspricht dabei einer Farbtemperatur von 6500 K und ist das ein Modell für indirektes Tageslicht. Für die theoretischen D-Leuchtkörper existiert jedoch aufgrund der gezackten Leistungsdichtespektren keine direkte technische Realisierung (Bild II.13). Es gibt jedoch einige Quellen, die als Tageslichtsimulatoren für die Kolorimetrie geeignet sind. Diese besitzen äquivalente Farbtemperaturen, die Struktur des Leistungsdichtespektrums weicht allerdings vom Standard ab. Die Druckindustrie verwendet als Bezug allgemein D50 und die Fotografie D55. Diese stellen Kompromisse zwischen den Beleuchtungen mit Kunstlicht (mit Wolframwendel) und Tageslicht dar. Für die menschliche Wahrnehmung hat das Licht einer modernen Wolframwendellampe mit ca. 3200 K eine gelbliche Erscheinung. Der theoretische Leuchtkörper D65 ist weitgehend farbneutral. Moderne blaue Leuchtstoffe von Computermonitoren und Fernsehgeräten sind im Bezug auf die menschliche Wahrnehmung wesentlich intensiver als rote oder grüne. Die Maximierung der Helligkeit der Monitore ging folglich mit einer Verringerung der Farbgenauigkeit einher, wobei Computermonitore mit einem Weißpunkt von 9300 K (D93) ungefähr doppelt so blau wie Tageslicht erscheinen. Dies ist allerdings für die Anwendung nicht problematisch, da sich das Auge an den dominanten Weißpunkt der Umgebung anzupassen vermag.

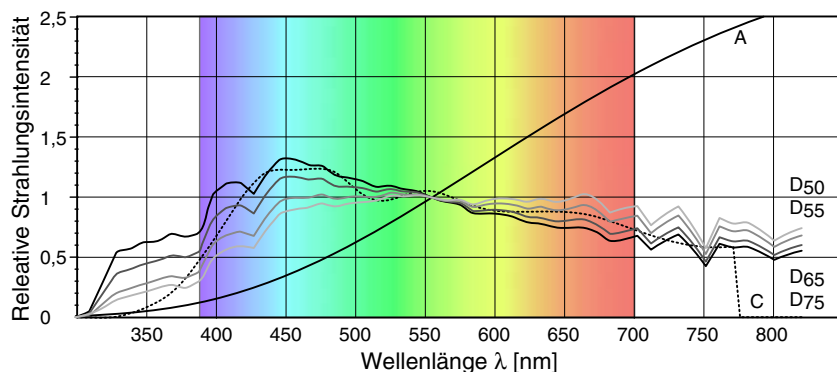


Bild II.13 Spektralverläufe der CIE-Leuchtkörper A, C und D50 bis D75

Ein isoliert betrachtetes Bild, z.B. ein Monitor oder ein in einen dunklen Raum projiziertes Dia, bewirkt beim Menschen seinen eigenen Weißpunkt. Auch größere Schwankungen der Farbtemperatur sind für die Betrachtung relativ unkritisch. Wird ein Bild jedoch unter Bezugnahme eines externen Weißpunktes betrachtet, so ergeben sich Wahrnehmungsunterschiede aus verschiedenen Weißpunkten. Das Farbempfinden wird dabei beeinträchtigt. Werden beispielsweise mit einem Kunstlichtfilm Aufnahmen bei Sonnenlicht erstellt, resultiert dies in bläulichen Farben. Vom Standpunkt der physikalischen Farbmessung ist dies jedoch kein Fehler. Der blaue Farbton ist dadurch begründet, daß der Kunstlichtfilm sehr blausensitiv ist, um die geringe kurzwellige Strahlung im Kunstlicht auszugleichen. Die Farbtemperatur wird dabei für die Wiedergabe zu höheren Farbtemperaturen verschoben. Die Referenzfarbtemperatur eines Kunstlichtfilms liegt bei ca. 3000 K und ermöglicht die farbneutrale Wiedergabe unter D55. Bei Belichtung mit Tageslicht (6000 K) ergeben sich folglich in den Blaubereich verschobene Farben. Tageslichtfilme besitzen dagegen eine Empfindlichkeit, die den Bedingungen bei Sonnenlicht angepaßt sind, um eine farbrichtige Wiedergabe von Tageslichtaufnahmen unter D55 zu ermöglichen. Als Schlußfolgerung läßt sich festhalten, daß Aufnahme und Wiedergabe abgestimmt werden müssen, um eine farbgetreue Reproduktion zu gewährleisten.

3.5 Konstanz der Wahrnehmung

Die spektrale Verteilung des natürlichen Lichts ändert sich in Abhängigkeit von Wetter, Tageszeit und geographischem Umfeld. Trotzdem vermag das Auge, unabhängig davon die Farbe von Objekten korrekt wahrzunehmen. Diese Fähigkeit des Auges, sich durch Umstimmung an veränderte Umgebungsbedingungen anzupassen und Farben auch bei veränderten Betrachtungsbedingungen und Eigenschaften der Lichtquelle wahrzunehmen, wird Konstanz der Wahrnehmung genannt. Konstanz der Wahrnehmung betrifft dabei sowohl die Reflektivität, die Farbe als Oberflächeneigenschaft, als auch die Albedo, die Helligkeit einer Oberfläche. Neben der Farbkonstanz gibt es z.B. auch eine Größenkonstanz oder eine Formkonstanz unter perspektivischer Verzerrung. Auch bei der Unterscheidung zwischen matten und glänzenden Oberflächen werden unabhängig von der Ausdehnung der Lichtquelle nur die Reflexionseigenschaften der Oberfläche wahrgenommen. Die wahrgenommene Farbe entspricht damit nicht einfach dem Spektrum des am Auge eintreffenden Lichtes, sondern hängt auch von den Spektren der Beleuchtung und den Oberflächeneigenschaften der beleuchteten Objekte ab. Vom Standpunkt der Informationsverarbeitung läßt sich dieses Problem der Farbwahrnehmung im wesentlichen darauf reduzieren, die Anteile von Lichtquelle und Reflektivität zu trennen.

3.6 Klassifikation der Wiedergabe von Farbe

Korrekte Farbwiedergabe bedeutet, daß eine aufgenommene Szene und deren Farben exakt so wiedergegeben werden, wie sie einem Beobachter auch im Original erscheinen. Bei dieser Forderung werden jedoch weder die Besonderheiten des menschlichen Sehsystems noch die technischen Randbedingungen berücksichtigt. Daher wurden Unterscheidungen der Farbwiedergabe vorgeschlagen [Hunt71], [Lang95]. Die Reproduktion läßt sich nach [Hunt95] in folgende Kategorien einteilen:

- exakte Reproduktion: Identität von Farbart, absoluter und relative Helligkeit

Bei der exakten Farbwiedergabe besitzt jede Körperfarbe bei der Reproduktion die gleiche Farbvalenz wie bei der Aufnahme vor der Kamera. Gleiche Farbvalenz fordert bekanntlich gleiche Farbart und Sättigung sowie gleiche Leuchtdichte. Die Leuchtdichte von Kathodenstrahlröhren hat jedoch technische Grenzen. Der Bereich der darstellbaren Leuchtdichten auf einem Farbfernseher ist viel geringer (max. $130 \text{ cd} \cdot \text{m}^{-2}$), als die bei der Aufnahme vorkommenden Leuchtdichten. Eine exakte Reproduktion ist somit technisch nicht realisierbar.

- kolorimetrischen Reproduktion: Identität von Farbart und relativer Helligkeit

Die farbmetrisch richtige Farbwiedergabe erzeugt bei der Wiedergabe für jede Körperfarbe die gleiche Farbart wie im Original. Die Leuchtdichte stimmt jedoch nur relativ zu einem Bezugsweiß überein. Bei einem Monitor liefert dies jedoch keine geeigneten Ergebnisse, da bei der Aufnahme oft verschiedene Lichtarten benutzt werden. Die Körperfarben erscheinen einem Beobachter im Original trotz unterschiedlicher Lichtarten in Folge der Umstimmung des Auges als gleich. Ein Beobachter vor dem Wiedergabegerät nimmt jedoch aufgrund des Wechsels der Lichtarten Unterschiede wahr. Auch andere Randbedingungen, wie Helligkeit und Farbe des Umfeldes sowie die Größe der Objekte, stimmen bei der Aufnahme und Wiedergabe nicht überein und beeinflussen die Farbempfindung.

- äquivalente Reproduktion: gleiches Aussehen von Farbart, absoluter und relativer Helligkeit
Bei der äquivalenten Farbwiedergabe sollen die Farben innerhalb einer Szene sowohl unabhängig von der Aufnahmelichtart als auch von anderen Bedingungen (Objektgröße und Umfeld) erscheinen. Dazu wäre es notwendig, daß bereits bei der Aufnahme alle Änderungen selbständig berücksichtigt und an die Wiedergabebedingungen angepaßt werden. Dies ist jedoch aufgrund der Komplexität nicht möglich.

- entsprechende Reproduktion: gleiches Aussehen von Farbart und relativer Helligkeit unter der Bedingung gleichen Helligkeitsniveaus

Bei der entsprechenden Reproduktion wird die Forderung der äquivalenten Reproduktion nach der Berücksichtigung der absoluten Helligkeit fallengelassen. Damit vereinfacht sich das Erfassungsproblem, da kameraseitig nur noch relative Helligkeiten zur Bestimmung der Farbarten berücksichtigt werden müssen. Die entsprechende Reproduktion ist noch Gegenstand aktueller Forschungen (CIE *Color Appearance Model*, CIECAM [Fair98]).

- Bevorzugte Reproduktion: angenehme Nachbildung von Gedächtnisfarben, indem die Identität des Aussehens vernachlässigt wird

Wenn die Forderung nur noch darin besteht, daß die Farben auf der Wiedergabeseite so erscheinen, wie sie dem Beobachter am besten gefallen, unabhängig davon, ob diese dann deutlich von der eigentlichen Farbe bei der Aufnahme abweichen, spricht man von der bevorzugten Farbwiedergabe.

Diese Unterscheidung der Arten der Farbwiedergabe kann z.B. zur Charakterisierung eines Wiedergabesystems verwendet werden. Beispielsweise sind für ein Farbfernsehsystem ungeeignet: die bevorzugte Farbwiedergabe (wegen der subjektiven Bewertung), die äquivalente/entsprechende Farbwiedergabe (auf Grund der Komplexität) wie auch die exakte Farbwiedergabe (wegen fehlender technischer Realisierbarkeit). Eine Charakterisierung ist somit nur über die kolorimetrische Farbwiedergabe möglich. Dabei ist eine Annäherung an die äquivalente Farbwiedergabe erreichbar, indem ein fester Bezugszustand definiert wird. Dieser Bezugszustand kann durch eine Normlichtart (nach EBU¹⁰ Normlichtart D65) gegeben sein, unter der die Farben auf dem Bildschirm erscheinen sollen. Bei der Wiedergabe müssen die Farbwerte zur Ansteuerung der Primärvalenzen des Bildschirms auf dieses Referenzweiß bezogen werden (Weißabgleich). In Kapitel III, Abschnitt 5 wird auf derartige Verfahren näher eingegangen. Bei der Aufnahme stellt sich somit die Forderung, daß eine Oberfläche der originalen Szene mit der Farbvalenz abgebildet werden muß, mit der sie bei der Beleuchtung unter Normlichtart D65 erscheinen würde. Dies kann einerseits dadurch geschehen, daß die Szene selbst mit D65 beleuchtet wird oder daß die Kamera eine abweichende Beleuchtung bei der Aufnahme kompensiert. Da für die theoretischen Leuchtkörper D65, D55, D50, etc. jedoch keine praktischen Lichtquellen existieren, bleibt nur die Kompensation während der Aufnahme. Dieses Problem deckt sich mit der Aufgabe zur Trennung von Reflektivität und Beleuchtungsanteil, die sich aus der Konstanz der Wahrnehmung ergibt.

10. EBU - European Broadcasting Union

Kapitel III

Farbmessung und -korrektur

柳は緑花は紅。

1 Farbmeßverfahren

Die Anwendung der Erkenntnisse der Funktion des Auges reicht von der allgemeinen Standardisierung und Vereinheitlichung von Aspekten der Licht- und Farbwahrnehmung bis zur konkreten Meßbarkeit von Farbe und Farbunterschieden. Damit wurde erreicht, daß Farben qualitativ und quantitativ vergleichbar sind. Mit der Farbkalibrierung durch Farbmeßgeräte ist auch eine Bewertung von Reproduktionsmedien (TV, LCD-Monitore, Druck, etc.) möglich. Die Meßbarkeit von Farbe beruht dabei auf Vereinbarungen, die in erster Linie bezwecken, die Meßbedingungen zu vereinheitlichen. Die wichtigste Bedingung ist dabei, daß alle Meßgeräte auf den gleichen spektralen Empfindlichkeitskurven basieren. Farbwahrnehmung ist jedoch keine absolute Sinnesgröße, sondern eine subjektive Empfindung, die von vielen Faktoren beeinflusst wird. In diesem Sinne haben Farbmessungen das Ziel, den visuellen Eindruck einer Farbe mit Farbmaßzahlen möglichst objektiv zu beschreiben.

Im Idealfall simulieren Farbmeßgeräte die Funktion des Auges. Bei stark vereinfachter Betrachtungsweise gilt für einen direkten Beobachter, daß eintreffendes Licht vom Auge in einen Sinneseindruck $s = f_{\text{Auge}}(l)$ umgewandelt wird. Die Menge der Sinneseindrücke $S \ni s$ wird dabei durch die Menge aller physikalisch möglichen Lichtspektren $L \ni l$ bestimmt. Bei einer Reproduktion wird das Licht zunächst mittels eines Sensors aufgenommen. Die Ausgabe des Sensors ist dabei ebenfalls eine Funktion des aufgenommenen Lichtes und liefert eine abstrakte Farbbeschreibung $m = f_{\text{Sensor}}(l)$. Die Sensorausgaben können durch Transformation in einen Farbraum übertragen werden, der beispielsweise zur Verarbeitung, Speicherung, Übertragung oder zur Wiedergabe der Farbinformationen geeignet ist. Der Wiedergabeprozess erzeugt Licht mit einer der Farbinformation entsprechenden spektralen Verteilung. Dieses Licht $wl = f_w(m)$ ruft im Auge eines menschlichen Betrachters einen Sinneseindruck hervor, der im Idealfall gleich dem Sinneseindruck ist, der vom originalen Licht verursacht wird. Somit gilt: $s = f_{\text{Auge}}(l) = f_{\text{Auge}}(wl)$. Mit der Wiedergabe- und Sensorfunktion ergibt sich somit: $s = f_{\text{Auge}}(l) = f_{\text{Auge}}(f_w(f_{\text{Sensor}}(l)))$. Die Funktion des Auges f_{Auge} stellt jedoch keine eindeutige Abbildung dar. Ebenso ist die Funktion des Sensors f_{Sensor} im allgemeinen keine eindeutige Abbildung. Eine Auflösung der Beziehung zur Ermittlung der zur farbrichtigen Aufnahme notwendigen Sensorfunktion f_{Sensor} ist somit nicht möglich. Wenn jedoch konstante Betrachtungsbedingungen vorausgesetzt werden, sind Sinneseindruck und Farbbeschreibung durch eine eindeutige Funktion f_x über $s = f_x(m)$ verbunden: Jede beschriebene Farbe erzeugt jeweils genau einen bestimmten Sinneseindruck.

Daraus folgt: $f_{\text{Auge}}(l) = s = f_x(m) = f_x(f_{\text{Sensor}}(l))$. Völlig unabhängig von der Wiedergabe ergibt sich somit, daß die Funktion des Sensors eindeutig mit der Funktion des Auges verbunden sein muß, um eine korrekte Reproduktion von Farbe zu gewährleisten.

Aus den Betrachtungen in Kapitel II, Abschnitt 2 läßt sich die Beziehung genauer spezifizieren: Sind zwei beliebige Primärfarbsysteme B_A und B_B mit den Primärfarben $\underline{b}_A(\lambda)$ und $\underline{b}_B(\lambda)$ gegeben, so besitzen diese je genau eine zugehörige Spektralwertfunktion $\underline{a}_A(\lambda)$ bzw. $\underline{a}_B(\lambda)$, die sich z.B. mit Farbmeßversuchen ermitteln lassen. Die Primärfarbsysteme sind dabei nach Kapitel II, Abschnitt 2.2.1 durch Lineartransformation verbunden. Die Transformationsmatrizen ergeben sich zu $\mathfrak{Z}_{12} = \int_{\lambda=0}^{\infty} \underline{a}_2(\lambda) \cdot \underline{b}_1(\lambda)^T \cdot d\lambda$, respektive $\mathfrak{Z}_{21} = \int_{\lambda=0}^{\infty} \underline{a}_1(\lambda) \cdot \underline{b}_2(\lambda)^T \cdot d\lambda$.

Da Aufgrund der Symmetrie der Farbtransformation auch $\mathfrak{Z}_{21} = \mathfrak{Z}_{12}^{-1}$ gilt, folgt:

$$\begin{aligned} \mathbf{E} &= \mathfrak{Z}_{12} \cdot \mathfrak{Z}_{12}^{-1} = \mathfrak{Z}_{12} \cdot \mathfrak{Z}_{21} = \mathfrak{Z}_{12} \cdot \int_{\lambda=0}^{\infty} \mathbf{a}_1(\lambda) \cdot \mathbf{b}_2(\lambda)^T \cdot d\lambda \\ &= \int_{\lambda=0}^{\infty} (\mathfrak{Z}_{12} \cdot \mathbf{a}_1(\lambda)) \cdot \mathbf{b}_2(\lambda)^T \cdot d\lambda \end{aligned} \quad (\text{III.1})$$

Aus dem Vergleich mit der Orthogonalitätsbedingung für Primärfarbsystem und Spektralwertfunktion nach Gleichung (II.20): $\mathbf{E} = \int_{\lambda=0}^{\infty} \mathbf{a}_2(\lambda) \cdot \mathbf{b}_2(\lambda)^T \cdot d\lambda$, folgt, daß $\mathbf{a}_2(\lambda) = \mathfrak{Z}_{12} \cdot \mathbf{a}_1(\lambda)$ sein muß. Dies bedeutet letztlich jedoch, daß die Spektralwertfunktionen aller Primärfarbsysteme, die zur Beschreibung der menschlichen Farbwahrnehmung geeignet sind, die Empfindlichkeitskurven des menschlichen Auges eingeschlossen, durch Linearkombination verbunden sind. Diese Beziehung wird auch Luther¹¹-Bedingung (oder auch Luther-Ives¹²-Bedingung) genannt [Luth27]. Während jede beliebige Kombination von unabhängigen Basisfarben als Primärfarbsystem geeignet ist, existiert nur eine einzige mögliche Spektralwertfunktion mit ihren Linearkombinationen für eine korrekte Farbaufnahme. Ein Sensor, der die menschlichen Farbwahrnehmung nachbildet, muß demzufolge eine spektrale Empfindlichkeit besitzen, die dem CIE Normalbeobachter oder einer Linearkombination davon entspricht.

1.1 Spektrale Farbmeßverfahren

Es existieren zwei prinzipiell verschiedene Methoden zur Messung von Farben. Die Grundidee der Spektralverfahren ist, daß jede Farbe als additive Mischung von einzelnen Spektralfarben aufgefaßt werden kann. Daraus folgernd beruht die Farbmessung mittels Spektralverfahren auf einer spektralen Zerlegung der zu messenden Strahlung in einzelne Anteile pro Wellenlänge und deren rechnerischen Verarbeitung mit den Spektralwertfunktionen des Normalbeobachters [Schäp93]. In der Praxis dient das Spektralphotometer zur Messung des Spektrums. Dabei wird das Spektrum in Abschnitte geteilt, deren Bandbreite 10 bis 20 nm beträgt. Jeder Abschnitt ergibt einen sogenannten Remissionswert. Das erste Ergebnis der Messung ist somit eine Reihe von Remissionswerten und ihre graphische Darstellung als Remissionskurve. Remissionswerte und Remissionskurve sind die vollständige Information der gemessenen Farbe. In der hier verwendeten Notation entsprechen die Remissionswerte der spektralen Strahlungsleistung Φ^λ und die Remissionskurve der spektralen Strahlungsleistungsdichte über der Wellenlänge $\varphi(\lambda)$. Zur Ermittlung der Normfarbwerte XYZ wird die Remissionskurve mit der Normspektralwertfunktion verknüpft. Dadurch ergibt sich explizit die Einhaltung der Lutherbedingung. Bezeichnen $x(\lambda)$, $y(\lambda)$, $z(\lambda)$ die Spektralwertfunktion des verwendeten Normbeobachters, so ergeben sich die Normfarbwerte X, Y, Z nach DIN5033 Teil 4 zu $X = \int_{\lambda=0}^{\infty} x(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$, $Y = \int_{\lambda=0}^{\infty} y(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$, $Z = \int_{\lambda=0}^{\infty} z(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$.

Für das Problem der Metamerie im Umgang mit Farbe hat sich das Spektralphotometer als nützliches Hilfsmittel bewährt. Am stärksten unterscheiden sich metamere Farben beim Wechsel von sehr verschiedenem Licht, etwa beim Wechsel von Tageslicht zum Kunstlicht. Der in der Norm DIN 6172 definierte Metamerie-Index wird deshalb vorzugsweise für den Wechsel von

11. Robert Luther (02.01.1868-14.04.1945) war von 1908 bis 1936 Direktor des Wissenschaftlich Photographischen Instituts der Technischen Hochschule Dresden und wandte als einer der ersten die Ergebnisse und Methoden der Physik und Chemie auf die Bearbeitung photographischer Probleme an.

12. F. E. Ives (1856-1937) erfand um 1892 ein Betrachtungsgerät zur Betrachtung von Farbausügen, das Chromoskop. Dieses basiert auf der Farbaufnahme mittels einer Kamera mit drei verschiedenfarbigen Filtern und subtraktiver Farbsynthese durch die zum jeweiligen Filter komplementären Farbstoffe.

den Tageslichtarten C, D50 und D65 zur Kunstlichtart A durch Messung mit einem Spektralphotometer bestimmt. Der entscheidende Nachteil spektraler Verfahren liegt jedoch in der notwendigen optischen Filterung zur Messung vieler schmalbandiger Spektralbereiche für jeden einzelnen Bildpunkt. Dies bedingt aufwendige optische Vorverarbeitung und langsame Verarbeitungsgeschwindigkeiten und spricht somit trotz der Genauigkeit gegen den breiten Einsatz dieser Verfahren in miniaturisierten Sensoren. Der Anwendung von Spektralphotometern ist daher auf die Messungen von Punkten oder Linien beschränkt. Matrizen von Sensorelementen zur direkten Aufnahme von Bildern sind auf dieser Grundlage praktisch nicht realisierbar.

1.2 Dreibereichsverfahren nach DIN5033

Das Dreibereichsverfahren bestimmt die Farbwerte der zu messenden Farbvalenz direkt. Dazu werden die Integrale von (mindestens) drei verschiedenen Filterungen der Strahlungsfunktion gemessen. Die verwendeten optischen Filter können dabei auch zeitlich nacheinander angewendet werden oder örtlich nebeneinander angeordnet sein. Diese Methoden besitzen jedoch spezifische Schwächen, die sich in einer Verringerung von zeitlichem bzw. örtlichem Auflösungsvermögen äußern.

Um eine korrekte Farbaufnahme zu gewährleisten, ist es erforderlich, daß die Empfindlichkeiten des Sensors an die spektralen Bewertungsfunktionen des gewählten Normbeobachters angepaßt sind. Sollen die Meßergebnisse unmittelbar die Normfarbwerte XYZ liefern, so müssen die Sensitivität des ungefilterten Sensors $s(\lambda)$ und die Transmission der drei Farbfilter $\tau_1(\lambda)$, $\tau_2(\lambda)$ und $\tau_3(\lambda)$ nach DIN5033 Teil 6 die Luther-Bedingung in folgender Form erfüllen: $x(\lambda) = s(\lambda) \cdot \tau_1(\lambda)$, $y(\lambda) = s(\lambda) \cdot \tau_2(\lambda)$, $z(\lambda) = s(\lambda) \cdot \tau_3(\lambda)$. Anstelle der Normspektralfunktionen $x(\lambda)$, $y(\lambda)$ und $z(\lambda)$ können auch beliebige Linearkombinationen dieser Funktionen treten. Die Meßergebnisse können anschließend durch lineare Transformation in das Normvalenzsystem überführt werden. Da praktisch verwendete Farbfilter die Luther-Bedingung jedoch nicht erfüllen, erreicht das Dreibereichsverfahren nicht die Meßgenauigkeit von Spektralphotometern. Wegen der einfachen Realisierbarkeit hat sich dieses Prinzip jedoch bei allen praktisch verfügbaren Sensoren für Farbbilder durchgesetzt.

Zur Verbesserung des Dreibereichsverfahrens ist die Verwendung von Teilmeßfiltern bekannt, die einen jeweils engeren Durchlaßbereich besitzen. Auch die Teilmeßfilter $\tau_1(\lambda) \dots \tau_n(\lambda)$ müssen die Lutherbedingung in folgender Form erfüllen:

$$\begin{aligned} x(\lambda) &= s(\lambda) \cdot (c_{x1} \tau_1(\lambda) + c_{x2} \tau_2(\lambda) + \dots + c_{xn} \tau_n(\lambda)) \\ y(\lambda) &= s(\lambda) \cdot (c_{y1} \tau_1(\lambda) + c_{y2} \tau_2(\lambda) + \dots + c_{yn} \tau_n(\lambda)) \\ z(\lambda) &= s(\lambda) \cdot (c_{z1} \tau_1(\lambda) + c_{z2} \tau_2(\lambda) + \dots + c_{zn} \tau_n(\lambda)) \end{aligned} \quad (\text{III.2})$$

Die notwendigen Konstanten $c_{x1} \dots c_{xn}$, $c_{y1} \dots c_{yn}$, $c_{z1} \dots c_{zn}$ sind entsprechend DIN5033 Teil 6 empirisch zu bestimmen.

2 Farbkorrektur

Die Zielsetzung bei der idealen Wiedergabe von Farbe ist, eine spektrale Intensitätsverteilung zu erzeugen, die eine bestimmte Farbempfindung im menschlichen Auge verursacht. Diese Farbempfindung soll dabei nicht von der ursprünglichen Farbe unterscheidbar sein. Im ersten Schritt der Farbproduktion, der Farbanalyse, wird daher im allgemeinen versucht, den

menschlichen Sehprozeß durch technische Farbsensoren nachzuahmen. In der Praxis hat sich dabei das Dreibereichsverfahren zur Aufnahme von Farbbildern durchgesetzt. In diesem Zusammenhang kann auch photographischer Film als Farbsensor betrachtet werden [Fisch97].

Trotz der vergleichsweise geringen Meßgenauigkeit im Vergleich zum Spektralphotometer ist besonders der einfache Aufbau des Dreibereichsverfahrens von praktischer Relevanz. Ähnlich dem menschlichen Auge werden drei spektrale Empfindlichkeiten verwendet, um die spektrale Intensitätsverteilung der abzubildenden Farbe in Sensorwerte umzuwandeln. In einem zweiten Schritt werden aus den Sensorwerten die Tristimuluswerte berechnet. Im abschließenden Schritt der Farbwiedergabe werden die Tristimuluswerte in Steuersignale für ein Ausgabegerät, beispielsweise in die RGB-Werte eines Monitors transformiert.

Die Transformation der Farbbeschreibung zur Wiedergabe und die Reproduktion der Farbe sind durch die technische Umsetzung festgelegt. Die CIE-Standardisierung erlaubt dabei, verschiedenste Geräte einheitlich zu charakterisieren. Die Aufnahme der Farbe und die Berechnung der Tristimuluswerte aus den Sensorwerten sind jedoch mit Schwierigkeiten verbunden. Dabei spielen fertigungstechnische Ungenauigkeiten der optischen Filter eine besondere Rolle. So sind optische Linsen nicht für alle spektralen Bereiche gleich durchlässig. Die notwendigen Spektralfilter lassen sich zudem nicht mit jeder beliebigen spektralen Charakteristik fertigen. Es existieren beispielsweise keine praktikablen Filter, die den CIE-Standardbeobachter vollständig nachbilden. Zusätzlich besitzen die photoaktiven Elemente der verschiedenen Sensoren unterschiedliche spektrale Empfindlichkeiten. CCD- und CMOS-Sensoren sind beispielsweise im roten und infraroten Bereich empfindlicher als das menschliche Auge. Das Maximum der menschlichen Wahrnehmung im roten Bereich liegt bei ca. 570 nm, während CCDs ein Maximum bei 700 nm besitzen und selbst noch im Infrarot-Bereich aktiv sind. Die Aufgabe der Farbkorrektur ist daher, die Farbvalenzen aus den Sensorausgaben zu ermitteln, die ein menschlicher Beobachter bei Betrachtung der Szene wahrnehmen würde.

Auch eine Änderung der Beleuchtung kann dazu führen, daß die Abstrahlung von Objekten zu unterschiedlicher Farbwahrnehmung führen. Die unterschiedlichen spektralen Verteilungen der Beleuchtung, kombiniert mit den Reflektanzen der Objekte, erzeugen unterschiedliche Beobachterantworten. Wird dies bei der Aufnahme nicht berücksichtigt, entstehen bei der Reproduktion Farbfehler in Form von unrealistischen Abschwächungen oder Betonungen einzelner Farbbereiche. Um eine farbmetrichtige Farbwiedergabe (siehe auch Kapitel II, Abschnitt 3.6) auf Bildschirmen zu ermöglichen, wurde festgelegt, daß ein Bild so aufgenommen werden muß, wie es einem Beobachter unter der Beleuchtung mit Normlichtart D65 erscheinen würde. Es existieren jedoch keine praktischen Lichtquellen für die theoretischen Leuchtkörper D65, D55, D50. Somit ist eine Kompensation des abweichenden Umgebungslichtes bei der Aufnahme erforderlich. Dies erfolgt durch Weißpunkt Korrektur oder Farbkonstanzalgorithmen.

2.1 Herkömmliche Methoden zur Farbkorrektur

Aus der Luther-Bedingung (Abschnitt 1) geht hervor, daß ein Photosensor die menschliche Farbwahrnehmung vollständig nachbilden muß, um eine fehlerfreie Reproduktion von Farben zu ermöglichen. Die erforderlichen spektralen Eigenschaften des Sensors ergeben sich dabei direkt aus den spektralen Empfindlichkeiten des CIE-Standardbeobachters oder aus einer Linearkombination davon. Abweichungen zwischen der implementierten Sensorcharakteristik und der

notwendigen spektralen Empfindlichkeit führen unweigerlich zu einer verfälschten Farabbildung. Aus den Betrachtungen zur Farbmessung läßt sich ableiten, daß keine Methode existiert, die Tristimuluswerte aus den Ausgangswerten von Sensoren zu berechnen, welche die Luther-Bedingung verletzen. In der Praxis wird dies dennoch häufig versucht.

Ein dabei oft verwendetes Verfahren ist die Korrektur mit Hilfe von Testbildern. Dazu werden Farbtafeln aufgenommen und die Ausgaben des Sensors mit den erwarteten Farbwerten verglichen. Die Methoden zur Anpassung reichen dabei von linearen Transformationen über nichtlineare Korrekturen bis zu neuronalen Netzen zur Farbklassifikation. Die linearen Abbildungen werden dabei oft durch 3×3 -Matrizen dargestellt und als Farbmatrixierung bezeichnet.

Auf dieser Basis ist eine Farbkorrektur jedoch nur unter Randbedingungen möglich, die für natürliche Bedingungen in der Regel nicht gegeben sind. Meist ist das Korrekturverfahren letztlich nur in der Lage, die getesteten Farben in identischer spektraler Zusammensetzung richtig einzuordnen. Zwischenfarben oder zu den getesteten Farben metamere Spektren führen dagegen zu falschen Ausgaben, die sich mehr oder weniger störend auswirken können. Da jedoch eine vergleichende Betrachtung von Original und Reproduktion meist nicht möglich ist, fallen diese Farbfehler nur in Extremfällen auf. Selbst größere Verschiebungen von reproduzierten Farben sind praktisch möglich. So wird sich in Ermangelung probater Möglichkeiten zur Farbkorrektur beispielsweise in der Farbphotographie vielfach auf die bevorzugte Reproduktion und die Nachbildung von Gedächtnisfarben zurückgezogen. Die farbmetrisch korrekte Wiedergabe von Farbe spielt dabei kaum noch eine Rolle.

Die Einschränkungen des wiedergebbaren Farbraumes aufgrund begrenzter Farbprimäre des entsprechenden Wiedergabemediums führen vielfach zu der Hypothese, daß bereits bei der Aufnahme der Bildinformationen ein eingeschränkter Farbraum und die Luther-Bedingung verletzende Sensorempfindlichkeiten verwendbar sind [Poy95]. Daß technische Sachzwänge eine in allen Bereichen korrekte Farbaufnahme behindern, steht dabei zwar außer Frage, jedoch kann durch analytische Betrachtung der Eigenschaften eines Systems zur Farbproduktion eine andere Aussage abgeleitet werden. Zum ersten ist die Entscheidung, ob eine Farbvalenz außerhalb oder innerhalb eines eingeschränkten Farbbereiches liegt, erst möglich, nachdem die Farbe korrekt aufgenommen wurde. Des weiteren kann es bei natürlichen Bildern durchaus vorkommen, daß Farben aufgenommen werden müssen, die im Wiedergabesystem nicht darstellbar sind. Diese Farben müssen für den eingeschränkten Wiedergabefarbraum so verschoben werden, daß keine sichtbaren Fehler durch z.B. Farbsprünge auftreten. Eine solche Korrektur ist aber nur dann möglich, wenn die formal nicht darstellbare Farbvalenz an sich korrekt aufgenommen wurde. Wird der Farbbereich jedoch von vornherein eingeschränkt, kann keine zuverlässige Korrektur stattfinden.

Weiterhin ist die Farbwahrnehmung von spektralen Empfindlichkeiten, welche die Luther-Bedingung verletzen, eine völlig andere im Vergleich zur menschlichen Wahrnehmung. Innerhalb des unendlichdimensionalen Raumes der Spektren S definiert die menschliche Wahrnehmung einen dreidimensionalen linearen Unterraum A , den Farbraum. Dabei existieren verschiedene Spektren $s \in S$, die auf denselben Vektor $\underline{a} \in A$, die entsprechende Farbe, abgebildet werden. Von der Luther-Bedingung abweichende Spektralwertfunktionen mit drei Koordinaten definieren ebenfalls einen dreidimensionalen Unterraum B im Spektralraum, der jedoch von Raum A verschieden ist. Im Farbraum B werden andere Spektren $s \in S$ gemeinsam auf bestimmte Vek-

toren $\underline{b} \in B$ abgebildet werden. Da die beiden Teilräume A und B nicht identisch und nicht linear abhängig sind, existiert kein analytischer Weg für eine Umrechnung. Ein Vektor \underline{b} in Raum B kann dabei mehreren Vektoren in Raum A entsprechen und umgekehrt. Eine Entscheidung, welche Vektoren in Raum B mit welchen Vektoren in Raum A korrespondieren, ist ohne explizite Kenntnis über den ursprünglichen Spektralraum nicht möglich.

Ein Ansatz zur Farbkorrektur, der diesem Gedankengang folgt, ist die Vektorraum-Approximation [Kön97]. Dabei wird im wesentlichen versucht, den Spektralraum aus dem Sensorraum B heraus zu schätzen, um anschließend den Raum des Normalbeobachters A zu berechnen. Allerdings kann kein analytischer Nachweis erbracht werden, ob eine derartige Schätzung erfolgreich sein kann. Unter stark eingeschränkten Versuchsbedingungen, die sich beispielsweise bei wenigen Testbildern ergeben, kann durchaus angenommen werden, daß eine derartige Schätzung der Spektren zu besseren Ergebnissen führt, als einfachere Methoden zur Farbkorrektur. Dies liegt darin begründet, daß in diesen Fällen die Struktur der Spektren stark begrenzt und eine Ermittlung der ursprünglichen Farben durch Klassifikation möglich ist. Für komplexere Umfeldbedingungen scheint das Verfahren jedoch nicht geeignet. Zu jedem Vektor in Raum B müßten dazu alle möglichen Spektren geprüft werden. Eine Entscheidung, welches der theoretisch unendlich vielen Spektren jedoch letztlich zu der spezifischen Sensorausgabe geführt hat, kann ohne Festlegung von strikten Randbedingungen nicht erfolgen.

3 Spektrale Rekonstruktion

3.1 Grundlagen

Allgemein betrachtet wird ein Sensor durch seine spezifische spektrale Empfindlichkeit $\underline{s}(\lambda)$, die von den Photoelementen und den Farbfiltern bestimmt ist, vollständig beschrieben. Die Ausgabe des Sensors bei linearem, integrierendem Aufbau sei definiert durch:

$$\underline{S} = \int_{\lambda=0}^{\infty} \underline{s}(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \quad (\text{III.3})$$

mit dem Leistungsdichtespektrum $\varphi(\lambda)$ des eintreffenden Lichtes. Entsprechend ist die Ausgabe des idealen Sensors mit der spektralen Sensitivität $\underline{a}(\lambda)$ definiert durch:

$$\underline{A} = \int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \quad (\text{III.4})$$

Die Transformation der Ausgaben des technischen Sensors \underline{S} zu Ausgaben des idealen Sensors \underline{A} soll durch Linearkombination erfolgen. Die beste Rekonstruktion der Farbvalenz \underline{A}' aus den Ausgaben des technischen Sensors \underline{S} ergibt sich dabei zu $\underline{A}'_i = \sum c_{ij} \cdot S_j$, beziehungsweise in Matrixschreibweise: $\underline{A}' = \underline{C} \cdot \underline{S}$.

Im klassischen Sensordesign werden \underline{S} und \underline{A} von gleicher Dimension gewählt und angestrebt, daß $\underline{C} = \underline{E}$ ist, um den Rechenaufwand zu minimieren. Damit ergibt sich formal $\underline{A}' = \underline{S}$ und auch $\underline{a}(\lambda) = \underline{s}(\lambda)$. Der Sensor muß demzufolge der Spektralwertfunktion des Zielfarbraumes vollständig entsprechen. Der Farbraum des Zielsystems ist jedoch in aller Regel ein RGB-Raum. Aus den Betrachtungen zu Farbprimären und Spektralwertfunktionen (siehe auch Kapitel II, Abschnitt 2.2) geht jedoch hervor, daß bei realen Farbprimären, wie sie in RGB-Räumen gegeben sind, immer irrealer Spektralwertfunktionen $\underline{a}(\lambda)$ auftreten. Die Empfindlichkeiten eines solchen zugehörigen Sensors weisen negative Anteile auf und sind somit technisch nicht

realisierbar. Die Empfindlichkeit $\mathfrak{s}(\lambda)$ eines technischen Sensors kann folglich prinzipiell nicht gleich der spektralen Sensitivität $\mathfrak{a}(\lambda)$ eines RGB-Raumes sein. Geht man von realisierbaren Spektralwertfunktionen aus, ist immer eine Transformation der Sensorausgaben in den Zielfarbraum notwendig. Die Annahme von $\mathfrak{C} = \mathbb{E}$ ist daher in der Praxis nicht zu erfüllen.

Ein standardisierter Farbraum mit realen Spektralwertfunktionen ist beispielsweise der CIE-XYZ-Raum mit dem Normalbeobachter. CIE-XYZ erlaubt dabei die Transformation in beliebige Zielfarbräume und ist daher zur Sensorkonstruktion geeignet. Es kann jedoch davon ausgegangen werden, daß eine vollständige Nachbildung des Normalbeobachters durch Farbfilter aufgrund von technologischen Beschränkungen nicht erreichbar ist (siehe auch Kapitel IV, Abschnitt 2). Ein erster Schritt der Farbkorrektur ist daher, eine Transformation \mathfrak{C} zu finden, mit der reale Sensorausgaben \mathfrak{S} in den CIE-XYZ-Raum umgerechnet werden können. Anschließend kann die Transformation in den gewünschten Wiedergabefarbraum erfolgen.

3.2 Die Methode der Spektralen Rekonstruktion

Die Grundlage für die Farbkorrektur mittels spektraler Rekonstruktion ist die Annahme, daß ein technischer Sensor durch Lineartransformation in einen idealen Sensor transformiert werden kann, der beispielsweise durch den CIE-Normalbeobachter gegeben ist. Da der CIE-Farbraum alle wahrnehmbaren Farben beschreiben kann, ist damit eine einheitliche Bewertung für verschiedene Sensoren über den gesamten Farbraum möglich. Die Transformation der Ausgaben des technischen Sensors \mathfrak{S} in die beste Rekonstruktion der Farbvalenz \mathfrak{A}' ergibt sich dabei zu:

$$\mathfrak{A}' = \mathfrak{C} \cdot \mathfrak{S} \quad (\text{III.5})$$

Aus der Gleichung (III.3) des Sensors folgt:

$$\mathfrak{A}' = \mathfrak{C} \cdot \int_{\lambda=0}^{\infty} \mathfrak{s}(\lambda) \cdot \varphi(\lambda) \cdot d\lambda = \int_{\lambda=0}^{\infty} [\mathfrak{C} \cdot \mathfrak{s}(\lambda)] \cdot \varphi(\lambda) \cdot d\lambda \quad (\text{III.6})$$

mit der spektralen Empfindlichkeit $\mathfrak{s}(\lambda)$ des technischen Sensors. Aus Vergleich von (III.6) mit (III.4) des idealen Sensors ergibt sich, daß die beste Rekonstruktion der spektralen Sensitivität des idealen Sensors $\mathfrak{a}'(\lambda) = \mathfrak{C} \cdot \mathfrak{s}(\lambda)$ sein muß, um (III.5) zu erfüllen.

Die Farbkorrektur durch Linearkombination der Sensorausgaben ist demzufolge möglich, falls die spektrale Sensitivität des idealen Sensors $\mathfrak{a}(\lambda)$ aus den technischen Sensitivitäten $\mathfrak{s}(\lambda)$ durch Linearkombination rekonstruierbar ist. Mathematisch entspricht dies einem Approximationsproblem. Die Ermittlung der optimalen Koeffizientenmatrix \mathfrak{C} kann dabei z.B. durch Minimierung des quadratischen Fehlers erfolgen. Der quadratische Fehler der Approximation ergibt sich dabei zu:

$$\mathbb{E} = \int_{\lambda=0}^{\infty} (\mathfrak{a}(\lambda) - \mathfrak{a}'(\lambda))^2 \cdot d\lambda = \int_{\lambda=0}^{\infty} (\mathfrak{a}(\lambda) - \mathfrak{C} \cdot \mathfrak{s}(\lambda))^2 \cdot d\lambda \quad (\text{III.7})$$

dessen Minimierung zu der allgemeinen Lösung¹³ nach [Bron87a] führt:

$$\int_{\lambda=0}^{\infty} \mathfrak{a}(\lambda) \cdot \mathfrak{s}(\lambda)^T \cdot d\lambda = \mathfrak{C} \cdot \int_{\lambda=0}^{\infty} \mathfrak{s}(\lambda) \cdot \mathfrak{s}(\lambda)^T \cdot d\lambda \quad (\text{III.8})$$

Die Umstellung nach der gesuchten Transformationsmatrix \mathfrak{C} ergibt:

$$\mathfrak{C} = \left(\int_{\lambda=0}^{\infty} \mathfrak{a}(\lambda) \cdot \mathfrak{s}(\lambda)^T \cdot d\lambda \right) \cdot \left(\int_{\lambda=0}^{\infty} \mathfrak{s}(\lambda) \cdot \mathfrak{s}(\lambda)^T \cdot d\lambda \right)^{-1} \quad (\text{III.9})$$

13. In Anhang A, Abschnitt 2.1 wird die allgemeine Lösung mit dem Verfahren der partieller Ableitung nach den Matrixelementen dargestellt.

$$\mathbf{C} = \int_{\lambda=0}^{\infty} \mathbf{a}(\lambda) \cdot \mathbf{b}(\lambda)^T \cdot d\lambda \quad \text{mit} \quad \mathbf{b}(\lambda) = \left(\int_{\lambda'=0}^{\infty} \mathbf{s}(\lambda') \cdot \mathbf{s}(\lambda')^T \cdot d\lambda' \right)^{-1} \cdot \mathbf{s}(\lambda) \quad (\text{III.10})$$

Die Funktionen $\mathbf{b}(\lambda)$ und $\mathbf{s}(\lambda)$ bilden dabei ein zusammengehöriges orthonormales System von Primärfarb- und Spektralwertfunktionen entsprechend Kapitel II, Abschnitt 2.

Die Transformation von Sensorwerten in Farbwerte des Zielfarbraumes ergibt sich somit zu:

$$\mathbf{A}' = \int_{\lambda=0}^{\infty} \mathbf{a}(\lambda) \cdot \left(\left(\int_{\lambda'=0}^{\infty} \mathbf{s}(\lambda') \cdot \mathbf{s}(\lambda')^T \cdot d\lambda' \right)^{-1} \cdot \mathbf{s}(\lambda) \right)^T \cdot d\lambda \cdot \mathbf{S} \quad (\text{III.11})$$

Aus dieser Beziehung lassen sich die Bedingungen ableiten, unter denen die Farbkorrektur vollständig ist. Dies ist zum ersten dadurch gegeben, wenn $\mathbf{a}(\lambda)$ und $\mathbf{s}(\lambda)$ linear abhängig sind. Das Integral in Gleichung (III.11) ergibt dabei die Abhängigkeit der spektralen Empfindlichkeiten und stellt somit einen formalen Beweis der Luther-Bedingung dar.

Neben der Luther-Bedingung läßt sich aus (III.11) noch eine zweite Bedingung ableiten, unter der eine vollständige Farbkorrektur möglich ist. Gleichung (III.11) stellt im Prinzip eine Sensorgleichung dar, bei der sich der Farbvektor \mathbf{A}' als Integral über das Produkt des Vektors der spektralen Sensitivität $\mathbf{a}(\lambda)$ mit einem Lichtspektrum ergibt. Das dabei verwendete Lichtspektrum ist die Rücktransformation der Sensorkoordinaten \mathbf{S} in den Spektralraum mit:

$$\varphi'(\lambda) = \left(\left(\int_{\lambda'=0}^{\infty} \mathbf{s}(\lambda') \cdot \mathbf{s}(\lambda')^T \cdot d\lambda' \right)^{-1} \cdot \mathbf{s}(\lambda) \right)^T \cdot \mathbf{S} \quad (\text{III.12})$$

Ist die Rücktransformation der Sensorwerte in den Spektralraum vollständig, so ist auch eine vollständige Farbkorrektur möglich.

Da jedoch bei einem praktischen Sensorsystem nur eine endliche Anzahl von Basisfunktionen $\mathbf{s}(\lambda)$ verfügbar ist, kann keine vollkommen fehlerfreie Rekonstruktion erreicht werden. Die Approximation ist allerdings um so genauer, je besser sich die Zielcharakteristik $\mathbf{a}(\lambda)$ mit den spektralen Empfindlichkeiten der technischen Sensorrealisierung nachbilden läßt. Dies ist entweder von der Qualität der Basisfunktionen oder von ihrer Anzahl abhängig. Mit vielen wenig angepaßten Funktionen können ähnlich gute Resultate erreicht werden wie mit wenigen gut passenden Funktionen. Ein Sensor mit genügender Anzahl von Farbkanälen kann somit einen idealen Sensor beliebig genau nachbilden, selbst wenn die einzelnen Kanäle fehlerhaft sind.

Anschaulich kann dies im Farbraum erklärt werden. Da eine direkte Umsetzung des unendlichdimensionalen Spektralraumes in den dreidimensionalen menschlichen Wahrnehmungsraum technisch schwierig ist, wird der Spektralraum vom Sensor zunächst in einen höher- aber endlichdimensionalen Vektorraum, den Sensorraum abgebildet. Diese Umwandlung ist in ihrer Qualität durch technologische Randbedingungen beschränkt. Anschließend erfolgt eine mathematische Transformation in den dreidimensionalen Wahrnehmungsraum, die exakt ist, wenn der darzustellende Wahrnehmungsraum ein Unterraum des Sensorraumes ist. Die praktische Umsetzung dieser beiden Aufgaben ist im Gegensatz zu einer direkten Transformation des Spektralraumes in den Wahrnehmungsraum von verhältnismäßig geringer Komplexität. Diese Betrachtung ermöglicht es, sich vom traditionellen Farbsensordesign zu lösen, das direkt auf der Tristimulustheorie basiert. Ein Sensor mit einer ausreichenden Anzahl von Farbkanälen kann einen idealen Sensor nachbilden, selbst wenn die einzelnen Farbkanäle fehlerbehaftet sind.

Die Wahl der gewünschten spektralen Empfindlichkeiten $\mathbf{a}(\lambda)$ des Zielsystems ist für die Rekonstruktion prinzipiell freigestellt. Es sind beliebige, beispielsweise auch praktisch nicht realisierbare Funktionen mit negativen Sensitivitäten rekonstruierbar.

Wird als Ziel der Rekonstruktion $\bar{a}(\lambda)$ der CIE-Normalbeobachter gewählt, so ist es sinnvoll, weitere Randbedingungen bei der Optimierung vorgeben, um der verzerrten Darstellung des CIE-XYZ-Raumes bezüglich der Wahrnehmung Rechnung zu tragen (siehe auch Kapitel II, Abschnitt 2.5). Dies kann beispielsweise über eine Skalierungsfunktion erfolgen, die den blauen und roten Spektralbereich gegenüber dem grünen verstärkt und so geringere Abweichungen der Kurven in diesen Bereichen ermöglicht.

3.3 Vergleich mit nichtlinearen Farbkorrekturen

In der Literatur sind auch nichtlineare Farbkorrekturmethoden beschrieben [Kön97]. Dem Ansatz nach wird eine nichtlineare Funktion f zur Berechnung der ideale Antwort A_1 aus den fehlerhaften Sensorausgaben S_1, \dots, S_n mit $A_1 = f(S_1, \dots, S_n)$ verwendet. Die nichtlineare Korrekturfunktion kann beispielsweise durch Polynomapproximation beschrieben werden:

$$\begin{aligned} A_1 &= f(S_1, \dots, S_n) = \\ &c_0 + c_{11}S_1 + c_{12}S_2 + \dots + c_{1n}S_n \\ &+ c_{211} \cdot S_1 \cdot S_1 + c_{222} \cdot S_2 \cdot S_2 + \dots + c_{2nn} \cdot S_n \cdot S_n \\ &+ c_{212} \cdot S_1 \cdot S_2 + \dots + c_{21n} \cdot S_1 \cdot S_n + \dots \end{aligned} \quad (\text{III.13})$$

Das Einsetzen der Sensorgleichung ergibt:

$$\begin{aligned} A_1 &= \int_{\lambda=0}^{\infty} a_1(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \\ &= c_0 + c_{11} \cdot \int_{\lambda=0}^{\infty} s_1(\lambda) \cdot \varphi(\lambda) \cdot d\lambda + c_{12} \cdot \int_{\lambda=0}^{\infty} s_2(\lambda) \cdot \varphi(\lambda) \cdot d\lambda + c_{1n} \cdot \int_{\lambda=0}^{\infty} s_n(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \\ &\quad + c_{211} \cdot \int_{\lambda=0}^{\infty} s_1(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \cdot \int_{\lambda=0}^{\infty} s_1(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \\ &\quad + c_{222} \cdot \int_{\lambda=0}^{\infty} s_2(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \cdot \int_{\lambda=0}^{\infty} s_2(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \\ &\quad + c_{2nn} \cdot \int_{\lambda=0}^{\infty} s_n(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \cdot \int_{\lambda=0}^{\infty} s_n(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \\ &\quad + c_{212} \cdot \int_{\lambda=0}^{\infty} s_1(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \cdot \int_{\lambda=0}^{\infty} s_2(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \\ &\quad + c_{21n} \cdot \int_{\lambda=0}^{\infty} s_1(\lambda) \cdot \varphi(\lambda) \cdot d\lambda \cdot \int_{\lambda=0}^{\infty} s_n(\lambda) \cdot \varphi(\lambda) \cdot d\lambda + \dots \end{aligned} \quad (\text{III.14})$$

Während der lineare Teil zur Approximation der spektralen Empfindlichkeit $a_1(\lambda)$ gemäß der Methode der spektralen Rekonstruktion verwendet werden kann, gilt dies nicht für die Terme höherer Ordnung, die aus Produkten von Sensorintegralen bestehen. Eine Optimierung der Koeffizienten ist bei diesen Termen abhängig vom eintreffenden Licht $\varphi(\lambda)$. Eine nicht-triviale Lösung kann somit nur für spezifische Eingangsspektren $\varphi(\lambda)$ angegeben werden. Da ein Farbkorrekturschema jedoch unter allgemeinen Bedingungen anwendbar sein muß, kann diese spezielle Falloptimierung nicht verwendet werden. Die Abhängigkeit der Optimierung vom eintreffenden Licht kann nur dann eliminiert werden, wenn die Koeffizienten der Terme höherer Ordnung gleich Null sind. Eine Farbkorrekturfunktion kann demnach prinzipiell keine Nichtlinearität aufweisen. Dies läßt sich auch aus der Funktionsweise der Rezeptorzellen des Auges ableiten.

4 Anwendung der Spektralen Rekonstruktion

Zur Demonstration der Farbkorrektur mittels spektraler Rekonstruktion sollen verschiedene Sensoren dienen, die z.T. im Rahmen anderer Projekte entwickelt wurden. Aus den Messungen von Photodioden einer 0,6 μm 2P3M und einer aktuellen 130 nm 1P6M CMOS-Technologie wurden die spektralen Empfindlichkeiten gewonnen und mit der Methode der spektralen Rekonstruktion Transformationsmatrizen ermittelt worden, die den praktischen Sensor in die jeweils beste Approximation eines idealen Sensors umrechnen. Als idealer Sensor dient dabei im allgemeinen der CIE-Normalbeobachter, da dieser einen standardisierten Vergleich verschiedener Sensoren ermöglicht. Zusätzlich werden mittels Simulation ermittelte hypothetische Empfindlichkeiten von neuen Sensorentwürfen betrachtet. Zu den Realisierungen dieser neuen Sensoren sei auf die nachfolgenden Kapitel verwiesen.

Als Bewertung der Approximation ist allgemein der quadratische Fehler nach Gleichung (III.7) verwendet worden und nicht, wie in der Kolorimetrie sonst üblich, der Farbabstand ΔE nach Anhang C, Abschnitt 2.6. Dies ist erforderlich, da der Farbabstand ΔE nur für ein spezielles Eingangslichtspektrum $\phi(\lambda)$ gültig ist und für jedes andere Spektrum unterschiedlich ausfällt, selbst wenn es sich um metamere Spektren handelt. Mit dem quadratische Fehler der Approximation der idealen Sensorempfindlichkeit wird dagegen eine vom Eingangslicht unabhängige Bewertung der Sensoren ermöglicht. Dieser Fehler kann zudem sehr einfach durch Gegenüberstellung von approximierter und geforderter Empfindlichkeit über der Wellenlängen veranschaulicht werden und erlaubt in dieser Darstellung zusätzlich eine qualitative Aussage zu Art und Größe der Abweichung in den verschiedenen Spektralbereichen. Der prinzipielle Ablauf der Farbkorrektur mittels spektraler Rekonstruktion ist aus Bild III.1 ersichtlich. Aus der Empfindlichkeit des Sensors $s(\lambda)$ und der gewünschten Zielcharakteristik $a(\lambda)$ ergibt sich die Transformation \mathbf{C} zur Umwandlung der Sensorausgabe in den gewünschten Zielfarbraum.

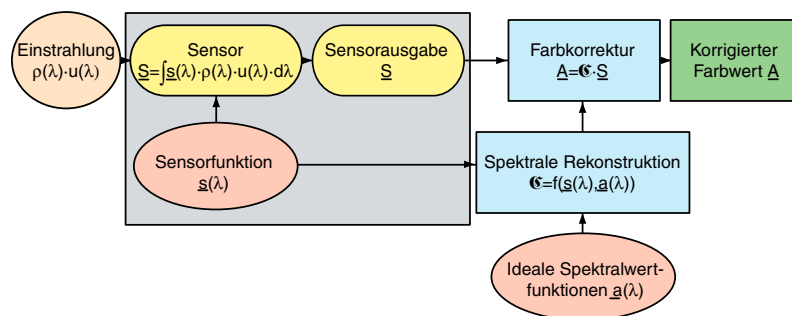


Bild III.1 Blockschaltbild für die Farbkorrektur mittels spektraler Rekonstruktion

4.1 3-Kanal-Sensor

In [RGH98] wird eine 1-Chip-CMOS-Kamera vorgestellt, die mit einem dreifarbigem Bayer-Mosaik (siehe Kapitel IV, Abschnitt 1.2.1) ausgestattet ist. Bild III.2 zeigt die gemessene spektrale Empfindlichkeit der drei Farbkanäle. Mit dem CIE-Normalbeobachter als Ziel ergibt sich eine Transformationsmatrix \mathbf{C} nach Bild III.2, die den kleinsten quadratischen Fehler zur Zielvorgabe gewährleistet. Die spektrale Empfindlichkeit des äquivalenten Sensors berechnet sich dabei zu $a'(\lambda) = \mathbf{C} \cdot s(\lambda)$ und ist ebenfalls in Bild III.2 dargestellt. Die Diskrepanz zwischen der Zielvorgabe und der besten Approximation ist klar erkennbar. Der zur Bewertung der Approximation verwendete quadratische Fehler fällt entsprechend der starken Abweichung der

Kurven, insbesondere im Rotbereich, sehr hoch aus. Die Ursachen liegen in der starken Infrarot-Empfindlichkeit des Sensors begründet. Weitere algorithmische Verbesserungen der Farbqualität sind nur mit Kenntnissen über Beleuchtung und Bildinhalt möglich, die unter praktischen Bedingungen jedoch in der Regel nicht vorhanden sind. Verbesserungen sind demzufolge nur durch Veränderung der spektralen Empfindlichkeit des Sensors erreichbar. Im einfachsten Fall kann dies durch ein externes Infrarotfilter geschehen. Die derart geänderte spektrale Empfindlichkeit des Systems führt zu einer Approximation des CIE-Normalbeobachters mit geringeren Abweichungen, wie in Bild III.3 zu sehen. Allerdings ist die Approximation noch immer nicht für qualitativ hochwertige Farbaufnahmen geeignet. Ohne weitere Veränderungen der Farbfilter sind bei diesem Sensorkonzept keine weiteren Verbesserungen möglich.

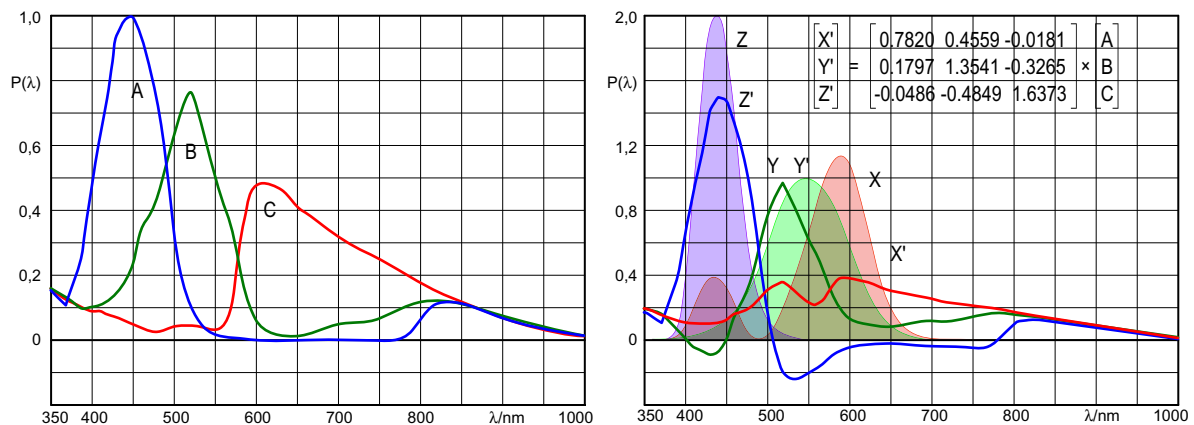


Bild III.2 Sensorfunktionen eines Drei-Kanal-Sensors und die CIE-Approximation

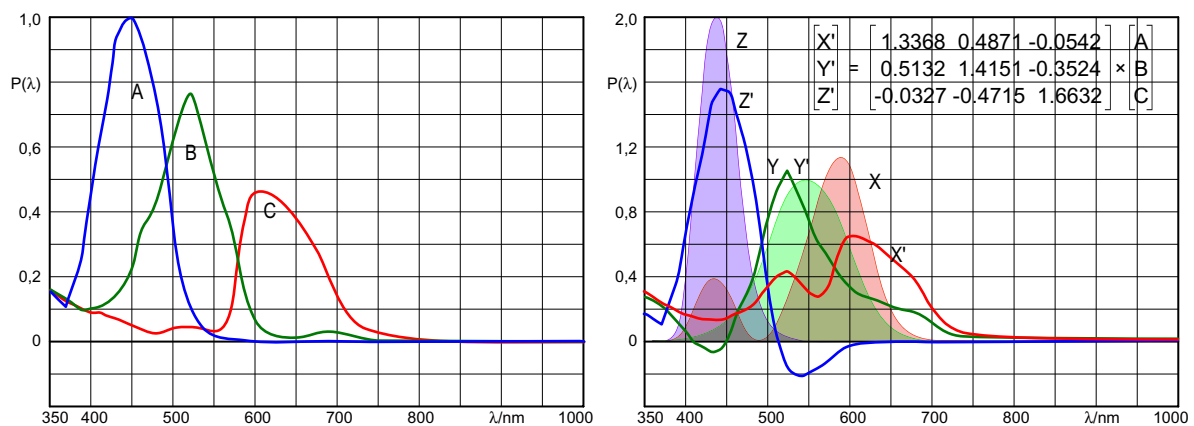


Bild III.3 Sensorfunktionen eines Drei-Kanal-Sensors mit IR-Filter und die CIE-Approximation

4.2 4-Kanal-Sensor

Die entscheidende Innovation zur Verbesserung der Farbkorrektur nach der Methode der spektralen Rekonstruktion ergibt sich aus der Möglichkeit, durch eine vergrößerte Anzahl von Sensorkanälen eine verbesserte Approximation der Zielfunktion zu erreichen.

Der vorliegende Fall des Sensors mit Bayer-Mosaik erlaubt die günstige Bereitstellung eines zusätzlichen Sensorkanals, ohne das Design selbst zu verändern. Da das Grundmuster des Bayer-Mosaiks aus einem roten, einem blauen und zwei grünen Pixeln besteht, ist es einfach möglich, eines der beiden grünen Pixel mit einer weiteren Farbe zu versehen. Im einfachsten Fall kann das Farbfilter auch ganz entfallen [HSS01]. Das entsprechende Pixel ist dann sensitiv für

den gesamten Wellenbereich. In Bild III.4 ist die spektrale Empfindlichkeit des Sensors und die beste Approximation des CIE-Normalbeobachters zu sehen. Im Vergleich zum 3-Kanal-Sensor ergibt sich die Fehlerreduktion im besonderen aufgrund der algorithmischen Unterdrückung der infraroten Störung. Zusätzlich zu den drei korrigierten Farbwerten kann beispielsweise auch ein Wert für den Infrarotanteil des eintreffenden Lichtes ermittelt werden. Dazu wurde eine spektrale Zielempfindlichkeit $i(\lambda)$ für den Infrarotbereich angenommen und mit den Sensorfunktionen die zugehörige Funktion $i'(\lambda)$ approximiert.

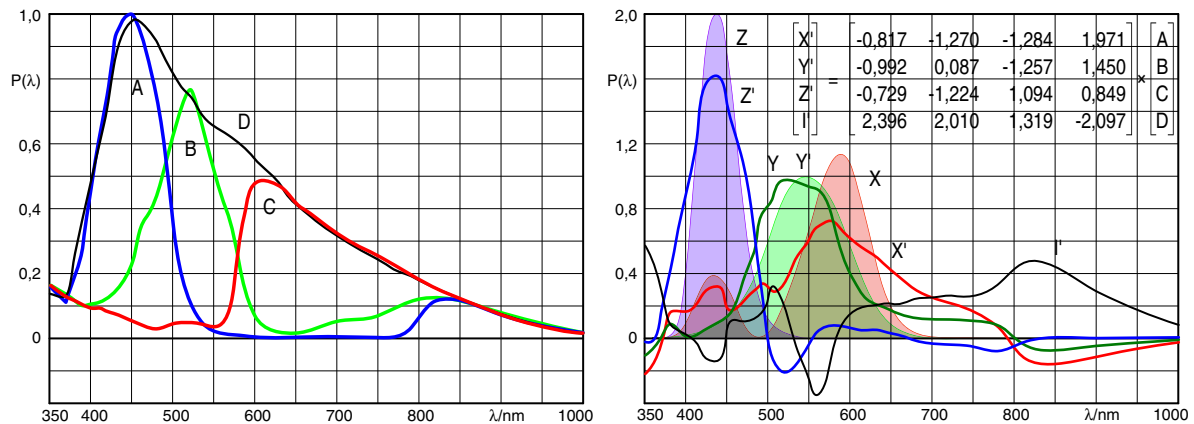


Bild III.4 Sensorfunktionen eines Vier-Kanal-Sensors und die CIE-Approximation

Für den praktischen Einsatz muß jedoch berücksichtigt werden, daß das Pixel ohne Farbfilter im Vergleich zu den Pixeln mit Farbfilter wesentlich mehr Strahlung empfängt und entsprechend stärker angesteuert wird. Für den Fall, daß alle Pixel schaltungstechnisch gleich dimensioniert sind, ergeben sich ein verringerter Dynamikbereich für die Pixel mit Farbfilter und folglich auch ein geringerer Signal-Rausch-Abstand. Der Aussteuerbereich kann jedoch z.B. durch eine teilweise Abschirmung des weißen Pixels angepaßt werden.

Sony hat am 16. Juli 2003 einen Farbsensor auf CCD-Basis mit Vier-Farb-Filter vorgestellt [Sony03]. Als zusätzlicher Farbfilter wird *Emerald* (Smaragdgrün, zwischen Cyan und Grün gelegen) verwendet. Laut Aussagen des Herstellers konnten die Farbfehler im Vergleich zu einem Sensor mit drei Farbkanälen halbiert werden. Diese Aussage läßt sich anhand des quadratischen Approximationsfehlers nach Gleichung (III.7) und der spektralen Rekonstruktion bestätigen. Bild III.5 zeigt die Verwendung eines solchen hypothetischen Farbfilters und die Approximation des CIE-Normalbeobachters. Ein ähnlicher Sensor ist in [HDBT97] beschrieben, der eine ähnliche vierte Farbe, *Teal* (Aquamarin, ebenfalls zwischen Blau und Grün) benutzt. Die Verwendung von zusätzlichen Farben durch weitere Vergrößerung des Filtermosaiks kann die Qualität der Farbaufnahme zwar verbessern, jedoch verringert sich dabei die Auflösung des Sensors und Fehler durch Farbmoiré-Effekte werden verstärkt (siehe auch Kapitel IV, Abschnitt 6.3). Deshalb wurden diese Möglichkeiten nicht weiter untersucht.

4.3 4-Kanal-CMOS-Sensor mit *Dual-Junction*-Photodioden

Die grundlegende Struktur der Zweikanal-Photozelle besteht aus einer *Dual-Junction*-Diode, die im nachfolgenden Kapitel detailliert beschrieben ist. Jedes Pixel mit *Dual-Junction*-Dioden besitzt zwei spektral verschiedene Kanäle. Um mit diesem zwei-kanaligen Sensor eine adäquate Farbaufnahme zu ermöglichen, ist ein zusätzlicher Absorptionsfilter erforderlich. Mit einem hypothetischen Zweifarben-Mosaik erhält der Sensor vier Farbkanäle, deren spektrale Empfindlich-

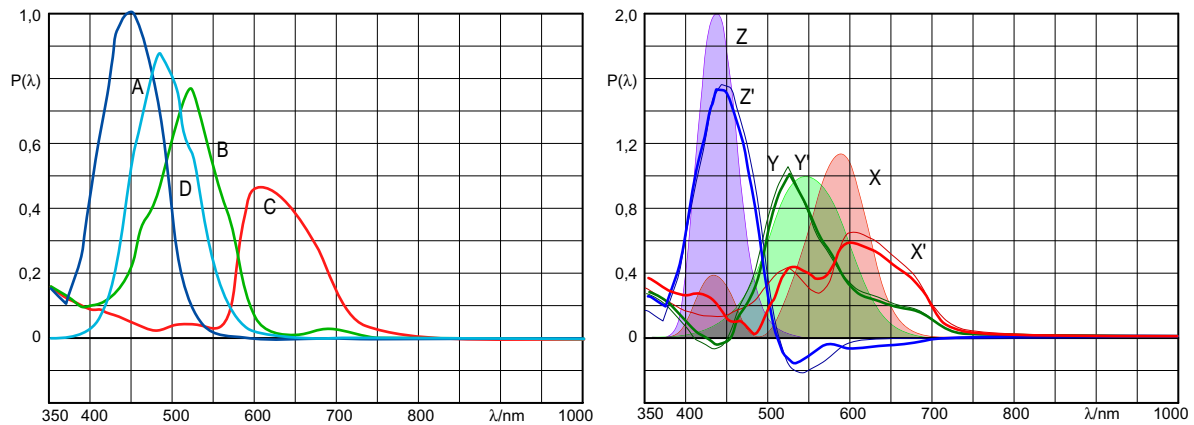


Bild III.5 Sensorfunktionen eines Vier-Kanal-Sensors mit Rot-Grün-Blau-Emerald- und IR-Filter mit CIE-Approximation (und der Drei-Kanal-Korrektur als Vergleich)

keit und die daraus gewonnene Empfindlichkeit des korrigierten Sensors in Bild III.6 dargestellt sind. Im Vergleich zum herkömmlichen Bayer-Mosaik verdoppelt sich dabei die optische Auflösung des Sensors, da nur zwei und nicht vier Nachbapixel für eine Interpolation der Farbwerte herangezogen werden müssen.

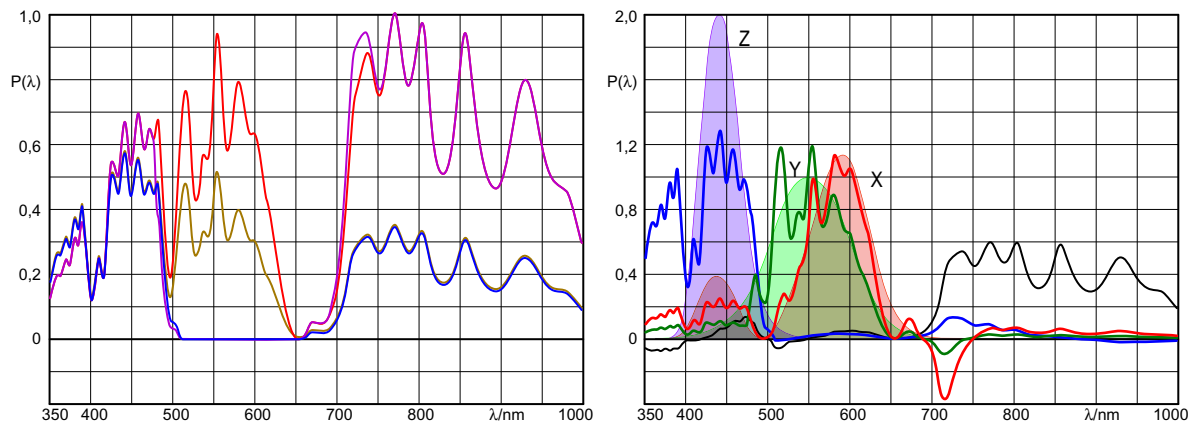


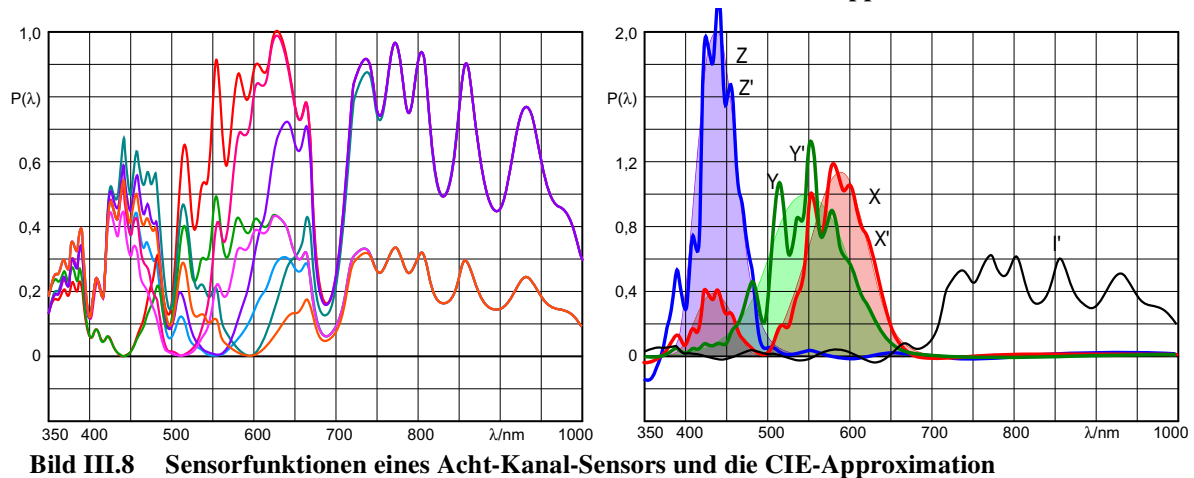
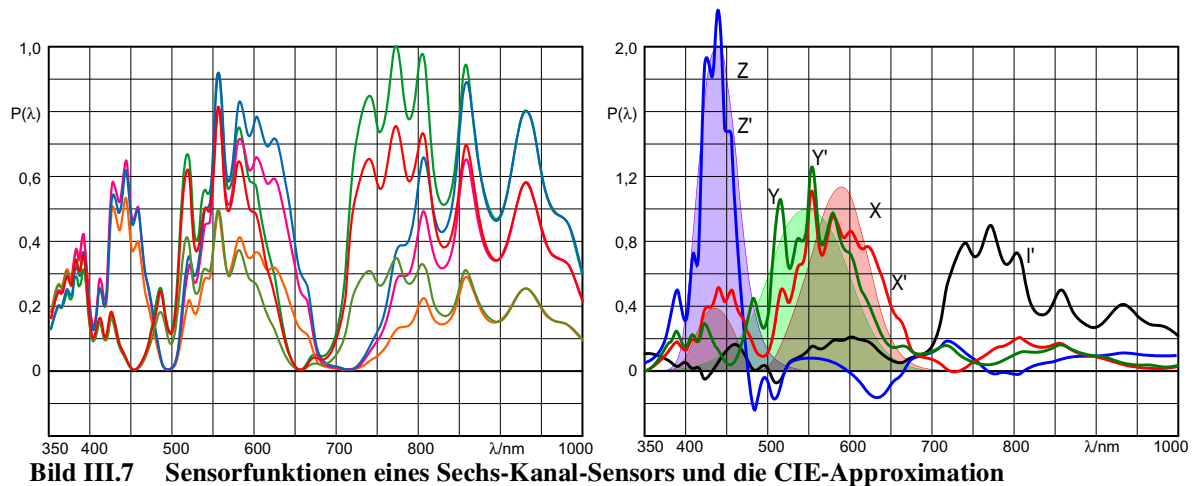
Bild III.6 Sensorfunktionen eines Vier-Kanal-Sensors und die CIE-Approximation

4.4 6-Kanal-CMOS-Sensor mit *Triple-Junction*-Photodioden

Bei *Triple-Well*/*BiCMOS*-Prozessen ist es möglich, die Struktur einer vertikalen Dreifachdiode zu realisieren. Das Basiselement besteht dabei aus drei vertikal angeordneten pn-Übergängen. Ohne Farbfilter entspricht die Struktur einem herkömmlichen Sensor mit drei Farbkanälen. Die erreichbare Farbqualität ist jedoch mangelhaft und wurde hier nicht weiter betrachtet. Mit einem Zweifarb-Mosaik lassen sich sechs spektral verschiedene Kanäle erzeugen, mit denen eine verbesserte Farbwiedergabe im Vergleich zum 4-Kanal-Sensor erreicht wird (Bild III.7).

4.5 8-Kanal-CMOS-Sensor mit *Dual-Junction*-Photodioden

Anstelle eines Zwei-Farb-Mosaiks kann für die Zweikanal-Photozelle nach Abschnitt 4.3 auch das modifizierte Bayer-Mosaik verwendet werden. Es ergeben sich dabei zwar keine Vorteile in Bezug auf die Auflösung im Vergleich zu herkömmlichen Farbsensoren, es stehen jedoch weitere Farbkanäle zur Verfügung, mit denen die Farbwiedergabe weiter verbessert werden kann. Die Kombination eines modifizierten Bayer-Mosaiks mit einer *Dual-Junction*-Photodiode ergibt 8 Farbkanäle, die eine nahezu ideale Farbaufnahme ermöglichen (Bild III.8).



4.6 Schlußfolgerungen

Aus den Darstellungen der Beispiele lassen sich Forderungen an die spektralen Sensorempfindlichkeiten ableiten. Allgemein kann mit einer größeren Anzahl von Sensorkanälen eine bessere Approximation der Zielcharakteristik erreicht werden. Da die Approximation der idealen Sensorantwort jedoch wegen der Linearkombination eine gewichtete Summe der Sensorkanäle darstellt, werden auch die Rauschterme der Sensors mit den entsprechenden Gewichten bewertet und addiert. Der Signal-Rausch-Abstand wird dabei besonders stark verringert, wenn die Zeilen der Transformationsmatrix Gewichte enthalten, die betragsmäßig ähnlich sind, jedoch unterschiedliche Vorzeichen aufweisen. Dies tritt beispielsweise auf, wenn sich die Empfindlichkeiten des Sensors überlappen und somit korreliert sind. Die Berechnung korrigierter Farbwerte aus stark korrelierten Sensorausgaben ist demzufolge besonders von Sensorrauschen betroffen [BH99]. So ist beispielsweise das resultierende Rauschen in Bild III.4 im Vergleich zu dem Ergebnis in Bild III.3 trotz der besseren Approximation mehr als doppelt so hoch. Die Summation des Rauschens tritt nicht auf, wenn die Transformationsmatrix nur einen dominierenden Term pro Zeile besitzt. Daraus folgt, daß die Einheitsmatrix die theoretisch günstigste Transformationsmatrix zur Korrektur darstellt, diese jedoch nach Abschnitt 3.1 praktisch nicht realisierbar ist. Es kann somit nur versucht werden, das Rauschen der Farbkorrektur und Farbraumtransformation durch geeignete Wahl der Farbfilter zu minimieren.

5 Weißpunktkorrektur und Farbkonstanz

5.1 Weißpunktkorrektur

Aus der Klassifikation der Farbwiedergabe wurde abgeleitet, daß ein Farbfernsehsystem nur durch die farbmetrisch richtige Farbwiedergabe möglich ist (siehe auch Kapitel II, Abschnitt 3.6). Eine Annäherung an die äquivalente Farbwiedergabe ist erreichbar, indem ein festes Bezugsweiß definiert wird. Dies geschieht in der Regel durch die Normlichtart D65. Für eine farbmetrisch richtige Farbwiedergabe ist eine originale Farbvalenz so aufzunehmen, wie sie einem Beobachter unter der Beleuchtung mit Normlichtart D65 erscheinen würde. Die Farbreize einer Szene sind bei nicht selbstleuchtenden Objekten durch die spektrale Reflektivität $\rho(\lambda)$ bestimmt. Wenn das Umgebungslicht von einer einzigen Lichtquelle mit dem Spektrum $u(\lambda)$ stammt, so wird die Ausgabe eines idealen Sensors bestimmt durch:

$$\underline{m}_u = \int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot u(\lambda) \cdot \rho(\lambda) \cdot d\lambda \quad (\text{III.15})$$

Gewünscht ist jedoch eine Ausgabe, die sich bei Beleuchtung mit D65 ergeben würde:

$$\underline{m}_{d65} = \int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot d65(\lambda) \cdot \rho(\lambda) \cdot d\lambda \quad (\text{III.16})$$

Ist das Spektrum des Umgebungslichtes bekannt, so kann diese Umrechnung direkt erfolgen, indem die Empfindlichkeit des Sensors $\underline{a}(\lambda)$ mit dem Faktor $d65(\lambda)/u(\lambda)$ gewichtet wird. Für dieses Vorgehen ist es jedoch erforderlich, daß $u(\lambda)$ über dem gesamten Definitionsbereich von Null verschieden ist. So ist es beispielsweise nicht möglich, eine monochromatische oder diskontinuierliche Szenenbeleuchtung in andere Lichtarten zu transformieren. Als künstliches Aufnahme­licht sind daher nur Tageslichtlampen mit kontinuierlichem Spektralverlauf geeignet, beispielsweise Glühlampen mit 5000-5500 K Lichttemperatur. Neon- und Halogenlicht besitzt dagegen Diskontinuitäten. Es fehlen ganze Spektralbereiche, die nicht ausgeglichen werden können. Spektrale Lücken sind auch bei sogenannten Halogen-Tageslichtlampen zu finden, die daher ebenfalls als Aufnahme­licht ungeeignet sind [Mar94].

Da ein Sensor jedoch nicht nur unter einer speziellen Lichtart arbeiten soll, muß eine Anpassung an die jeweils vorherrschende Beleuchtung der Szene erfolgen. Die Anpassung an die Aufnahmebeleuchtung durch Korrektur des Unbunt­punktes wird dabei als Weißabgleich bezeichnet. Die Realisierung kann sowohl schaltungstechnisch durch unterschiedlich empfindliche Sensoren als auch mit speziellen Filtern erfolgen. Da jedoch beiden Möglichkeiten sehr kostenintensiv sind, wird die Anpassung der spektralen Empfindlichkeit an die sich wechselnde Beleuchtung in der Praxis über einen Zwischenschritt vorgenommen. Die Sensorausgaben werden zunächst in einen RGB-Farbraum transformiert. Die Verstärkungsfaktoren k_r , k_g und k_b der Primärfarben Rot, Grün und Blau werden nun so ausbalanciert, daß sich für eine mattweiße Vorlage mit $\rho(\lambda) = 1$ die Farbwerte R, G, B zu je 100% ergeben. Werden diese Farbwerte anschließend an das Bildwiedergabesystem übertragen, ergibt sich dort die Normlichtart D65, die als Bezugslichtart von der EBU¹⁴ vorgeschrieben wurde. Die Verstärkungsfaktoren werden anschließend auf den Grünkanal normiert. Der Weißabgleich erfolgt somit nur über die Veränderung der Verstärkungen des Rot- und Blaukanals. Damit werden die spektralen Empfindlichkeiten jedoch nur skaliert, nicht aber ihr spektraler Verlauf verändert. Andere

14. EBU - European Broadcasting Union

Punkte im Farbraum außer dem Unbuntpunkt werden zum Abgleich nicht herangezogen. Dies kann dazu führen, daß zwar weiße Objekte in der Szene korrekt wiedergegeben werden können, andere Farben jedoch weiterhin fehlerhaft sind.

5.2 Farbkonstanz

Neben dem Weißabgleich existieren Methoden, die den vollständigen Farbraum betrachten, um Änderungen der Umgebungsbeleuchtung quantitativ zu erfassen. In [Haf98] wird ein solcher Farbkonstanzalgorithmus vorgestellt, der hier stellvertretend veranschaulicht werden soll. Die durch den Sensor wahrgenommene Farbe eines Bildpunktes setzt sich allgemein aus der spektralen Leistungsdichte der beleuchtenden Strahlung $u(\lambda)$, dem spektralen Reflexionsfaktor der Oberfläche $\rho(\lambda)$ und der spektralen Empfindlichkeit des Sensors $a(\lambda)$ zusammen. Wird die spektrale Verteilung des Umgebungslichts durch eine Anzahl von Basisfunktionen $b_i(\lambda)$ angenähert, so kann die Leistungsdichte des Umgebungslichtes durch die Summe der Basisfunktionen dargestellt werden: $u(\lambda) = \sum_i c_i \cdot x_i(\lambda) = \underline{c}^T \cdot \underline{b}(\lambda)$, mit den Koeffizienten c_i und deren Vektordarstellung \underline{c} . Die Ausgabe des Sensors ergibt sich dann entsprechend zu:

$$\underline{m} = \int_{\lambda=0}^{\infty} a(\lambda) \cdot [\underline{c}^T \cdot \underline{b}(\lambda)] \cdot \rho(\lambda) \cdot d\lambda = \int_{\lambda=0}^{\infty} a(\lambda) \cdot \underline{b}(\lambda)^T \cdot \rho(\lambda) \cdot d\lambda \cdot \underline{c} = \underline{\mathfrak{Z}} \cdot \underline{c} \quad (\text{III.17})$$

Ist das Umgebungslicht eine Standardbeleuchtung $u'(\lambda)$, ergibt sich $u'(\lambda) = \underline{c}'^T \cdot \underline{b}(\lambda)$ und die Sensorausgabe ist verkürzt: $\underline{m}' = \underline{\mathfrak{Z}} \cdot \underline{c}'$. Die Transformation zwischen \underline{m} und \underline{m}' sei durch eine unbekannte lineare Transformation $\underline{\mathfrak{X}}$ zu $\underline{m}' = \underline{\mathfrak{X}} \cdot \underline{m}$ bestimmt. Da sich $\underline{\mathfrak{Z}}$ aus dem spektralen Reflexionsfaktor des Oberflächenmaterials $r(\lambda)$ sowie der spektralen Empfindlichkeit des Sensors $a(\lambda)$ zusammensetzt, ändert sich $\underline{\mathfrak{Z}}$ nur mit dem betrachteten Objekt. Werden weiterhin \underline{c} und \underline{c}' als szenenunabhängig vorausgesetzt, kann $\underline{\mathfrak{Z}}$ aus der Gleichung eliminiert werden, um die Transformation $\underline{\mathfrak{X}}$ zu ermitteln. Die Matrix $\underline{\mathfrak{Z}}$ kann jedoch nur dann entfernt werden, wenn sowohl $\underline{\mathfrak{Z}}$ als auch $\underline{\mathfrak{X}}$ Diagonalmatrizen sind:

$$\underline{\mathfrak{Z}} \cdot \underline{c}' = \underline{\mathfrak{X}} \cdot \underline{\mathfrak{Z}} \cdot \underline{c} \quad \underline{c}' = \underline{\mathfrak{Z}}^{-1} \cdot \underline{\mathfrak{X}} \cdot \underline{\mathfrak{Z}} \cdot \underline{c} = \underline{\mathfrak{X}} \cdot \underline{c} \quad (\text{III.18})$$

Für die Bestimmung des Einflusses einer Beleuchtungsänderung auf die Farben in der Szene ist in diesem Fall das Wissen über die globale Farbänderung $\underline{\mathfrak{X}}$ ausreichend. Zur Berechnung der Transformationsmatrix $\underline{\mathfrak{X}}$ werden in der Praxis Vergleichsbilder verwendet, die dieselbe Szene unter den unterschiedliche Beleuchtungen $u(\lambda)$ und $u'(\lambda)$ zeigen. Aus den Bildern werden Referenzfarben extrahiert und der jeweilige Transformationsvektor bestimmt. Dies führt zu einem überbestimmten Gleichungssystem, dessen Minimierung die gewünschte Lösung $\underline{\mathfrak{X}}$ ergibt, mit der zu jeder gemessenen Farbvalenz die entsprechende Farbvalenz unter der Standardbeleuchtung mit $\underline{m}' = \underline{\mathfrak{X}} \cdot \underline{m}$ berechnet werden kann. Damit die Matrix $\underline{\mathfrak{Z}}$ Diagonalgestalt hat, müssen die Sensorempfindlichkeit $a(\lambda)$ und die Basisfunktion des Umgebungslichtes $b(\lambda)$ orthogonal sein. Diese wesentliche Voraussetzung kann jedoch nach [Haf98] in der Praxis nicht erfüllt werden. Wenn $\underline{\mathfrak{Z}}$ keine Diagonalmatrix ist, variiert die Transformation zwischen $u(\lambda)$ und $u'(\lambda)$ je nach betrachtetem Objekt. Eine lineare Transformation $\underline{\mathfrak{X}}$ zur globalen Kompensation der Beleuchtungsänderung zwischen zwei Bildern kann dann nicht angegeben werden. Es existiert zwar eine Reihe von Erweiterungen dieses Ansatzes, die von allgemeineren Voraussetzungen ausgehen, allerdings auch rechenintensiver sind. Die Gemeinsamkeiten dieser Verfahren liegen jedoch darin, daß die Informationen über die verwendete Beleuchtung aus dreikanaligen Farb-

bildern extrahiert werden und eine Bestimmung der spektralen Verteilung des Umgebungslichts daher nur unzureichend erfolgen kann.

5.3 Weißabgleich und Farbkonstanz mit Multikanal-Sensoren

Die Möglichkeit, beliebige Sensorfunktionen mittels der spektralen Rekonstruktion nachzubilden, eröffnet auch bei Weißabgleich und Farbkonstanz neue Wege. Im wesentlichen benötigen diese Anwendungen die Trennung von Beleuchtung und Reflektanz der Objekte. Eine Farbvalenz ist bei einer spektralen Leistungsdichte des Umgebungslichtes $u(\lambda)$ und der spektralen Reflektanz der Oberfläche $\rho(\lambda)$ gegeben durch $\underline{m}_u = \int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot u(\lambda) \cdot \rho(\lambda) \cdot d\lambda$, wobei $\underline{a}(\lambda)$ die spektrale Empfindlichkeit eines idealen Sensors mit z.B. CIE-Charakteristik darstellen soll. Bei einem modifizierten Idealsensor mit der spektralen Empfindlichkeit $\underline{a}'(\lambda) = \underline{a}(\lambda) / u(\lambda)$ ergeben sich die Farbvalenzen zu:

$$\underline{m}'_u = \int_{\lambda=0}^{\infty} \underline{a}'(\lambda) \cdot u(\lambda) \cdot \rho(\lambda) \cdot d\lambda = \int_{\lambda=0}^{\infty} \frac{\underline{a}(\lambda)}{u(\lambda)} \cdot u(\lambda) \cdot \rho(\lambda) \cdot d\lambda = \int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \rho(\lambda) \cdot d\lambda \quad (\text{III.19})$$

Die zu bestimmenden Farbvalenz wird folglich nicht mehr vom Umgebungslicht beeinflusst.

Durch die Methode der spektralen Rekonstruktion nach Abschnitt 3.2 kann eine ideale Sensorfunktion durch eine reale Sensorfunktion $\underline{s}(\lambda)$ mit $\underline{a}(\lambda) = \underline{\mathbf{G}} \cdot \underline{s}(\lambda)$ approximiert werden. Dies gilt auch für die modifizierte Empfindlichkeit $\underline{a}'(\lambda)$ und ergibt eine Korrekturmatrix $\underline{\mathbf{G}}_u$, mit der die Zielfunktion $\underline{a}'(\lambda) = \underline{a}(\lambda) / u(\lambda) = \underline{\mathbf{G}}_u \cdot \underline{s}(\lambda)$ angenähert wird. Eine Änderung der Sensoreigenschaften ist somit allein durch Modifikation der Korrekturmatrix möglich, wohingegen bei klassischen Sensoren spezielle Filter und Eingriffe in das optische System notwendig sind. Um eine entsprechende Genauigkeit zu gewährleisten, sind jedoch ausreichend viele Sensorkanäle notwendig. Auch die Umrechnung der Farben in andere Lichtarten sind mit dieser Methode möglich. Dazu wird die Sensorfunktion mit der spektralen Leistungsdichte des gewünschten Umgebungslichtes gewichtet. Für Normlicht D65 kann eine entsprechende Sensorfunktion demnach mit $\underline{a}'(\lambda) = \underline{a}(\lambda) \cdot d65(\lambda) / u(\lambda) = (\underline{\mathbf{G}}_{u, d65} \cdot \underline{s}(\lambda))$ erhalten werden. Die mit dem Sensor aufgenommenen Farbvalenzen ergeben sich dann so, als würden die Szene mit Normlicht D65 beleuchtet.

Durch die Möglichkeit der algorithmischen Änderung der Sensorparameter wird auch eine dynamische Anpassung an wechselnde Beleuchtungsverhältnisse möglich. Nach dem Ansatz aus Abschnitt 5.2 ist ausgehend von den Sensorfunktionen $\underline{s}(\lambda)$ eine direkte Approximation des Umgebungslichtes $u(\lambda)$ realisierbar, um eine Kompensation von unbekannten Umgebungsbedingungen zu ermöglichen. Allgemein wird dies auch als automatischer Weißabgleich bezeichnet. Wenn ein mattweißes Objekt mit der Reflektivität $\rho(\lambda) = 1$ beleuchtet wird, ergibt sich eine Sensorausgabe $\underline{m}_w = \int_{\lambda=0}^{\infty} \underline{s}(\lambda) \cdot u(\lambda) \cdot d\lambda$. Um aus diesen Werten \underline{m}_w eine Näherung des Spektrums $u(\lambda)$ mit $u'(\lambda) = \underline{b}(\lambda)^T \cdot \underline{m}_w$ zu berechnen, sind Rekonstruktionsfunktionen $\underline{b}(\lambda)$ notwendig. Für diese Funktionen wurde in Kapitel II, Abschnitt 2.2.1 aus dem allgemeinen mathematischen Modell eines Sensors eine Bedingung abgeleitet, unter welcher ein Eingangsspektrum $\phi(\lambda)$ aus den zugehörigen Ausgabewerten eines Sensors \underline{m} mit $\phi'(\lambda) = \underline{m}^T \cdot \underline{b}(\lambda)$ rekonstruiert werden kann. Diese lautet allgemein:

$$\underline{b}(\lambda) = \left(\int_{\lambda=0}^{\infty} \underline{c}(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{c}(\lambda) \quad (\text{III.20})$$

mit der vollständigen Beschreibung der Spektren $\phi'(\lambda) = \underline{\Phi}^T \cdot \underline{c}(\lambda)$ und $\phi(\lambda) = \underline{\Phi} \cdot \underline{c}(\lambda)$

durch Basisfunktionen $\underline{s}(\lambda)$. Für $\underline{s}(\lambda) = \underline{s}(\lambda)$ ergibt sich nach Gleichung (II.13):

$$\underline{b}(\lambda) = \left(\int_{\lambda=0}^{\infty} \underline{s}(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{s}(\lambda) \quad (\text{III.21})$$

Da die resultierenden Rekonstruktionsfunktionen jedoch Linearkombinationen der Sensorfunktionen sind, ist auch die Nachbildung des Eingangslichtspektrums nur eine Linearkombinationen der Sensorfunktionen. Eine ideale Rekonstruktion des Eingangslichtspektrums ist somit für praktisch relevante Anwendungen mit begrenzter Anzahl von Basisfunktionen nicht möglich. Für die Anwendung in einem Farbkonstanzalgorithmus kann diese Approximation jedoch ausreichend sein. Allerdings sind die Rekonstruktionsfunktionen, die sich direkt aus Gleichung (III.21) ergeben, nur bedingt geeignet. Da die zwangsläufig auftretenden Meßfehler erheblichen Einfluß auf den Verlauf der Funktionen $\underline{b}(\lambda)$ haben, ist auch die daraus gebildete Approximation der spektralen Leistungsdichte des Umgebungslichtes stark von statistischen Schwankungen betroffen und kann somit allgemein keine direkte Verwendung in Gleichung (III.19) finden. Zur Aufbereitung bietet sich die Transformation der Sensorausgaben in einen anderen Farbraum an, dessen Primärfarbfunktion geeignet festzulegen ist. Das Schema soll im folgenden am Beispiel des 8-Kanal-Sensors nach Bild III.8 illustriert werden.

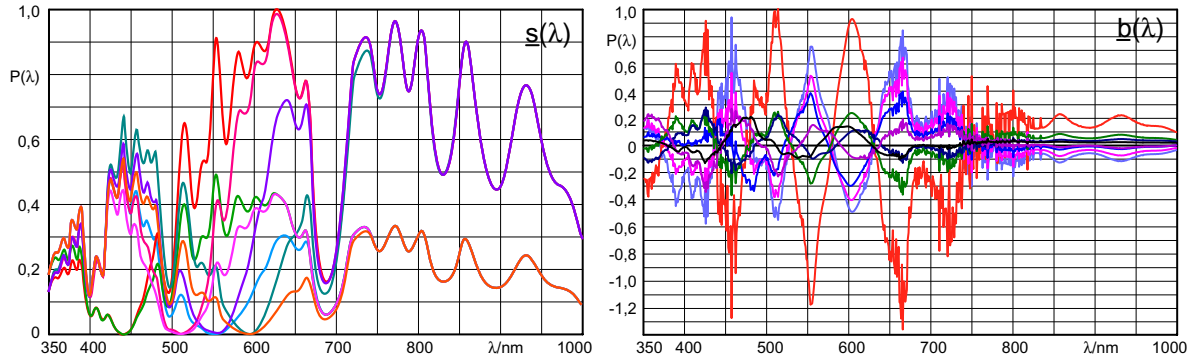


Bild III.9 Spektralwertfunktionen (links) und Primärfarben (rechts) des 8-Kanal-Sensors

Zur Veranschaulichung sind die Kurven der spektralen Empfindlichkeiten der einzelnen Sensorkanäle erforderlich, die in Bild III.9 zu sehen sind. Die Spektralwertfunktion $\underline{s}(\lambda)$ des Sensors ist aufgrund von technologischen Randbedingungen festgelegt. Nach Gleichung (III.21) ergibt sich mit $\underline{b}(\lambda) = \left(\int_{\lambda=0}^{\infty} \underline{s}(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{s}(\lambda)$ das entsprechend zugehörige Primärfarbsystem des Sensors. Das Farbsystem ist durch $\underline{s}(\lambda)$ und $\underline{b}(\lambda)$ vollständig bestimmt. Die Sensorausgabe sei durch $\underline{m}_s = \int_{\lambda=0}^{\infty} \underline{s}(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$ gegeben.

Zur Approximation des originalen, auf den Sensor wirkenden Spektrums kann ein beliebiges Primärfarbsystem $\underline{b}_x(\lambda)$ festgelegt werden. Für das Beispiel sollen die Kurven nach Bild III.10 verwendet werden. Die äquivalenten Spektralwertfunktionen des Systems berechnen sich nach Gleichung (II.14) zu $\underline{s}_x(\lambda) = \left(\int_{\lambda=0}^{\infty} \underline{b}_x(\lambda) \cdot \underline{b}_x(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{b}_x(\lambda)$ aus den gewählten Primärfarben. Die entsprechende Ausgabe des theoretischen Sensors ist dann bestimmt durch $\underline{m}_x = \int_{\lambda=0}^{\infty} \underline{s}_x(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$. Setzt man $\varphi(\lambda) = \varphi'(\lambda) = \underline{b}(\lambda)^T \cdot \underline{m}_s$ und obige Gleichungen ineinander ein, ergibt sich die Umrechnung von \underline{m}_s nach \underline{m}_x zu:

$$\underline{m}_x = \int_{\lambda=0}^{\infty} \underline{s}_x(\lambda) \cdot \underline{b}(\lambda)^T \cdot \underline{m}_s \cdot d\lambda$$

$$\underline{m}_x = \int_{\lambda=0}^{\infty} \left(\int_{\lambda'=0}^{\infty} \underline{b}_x(\lambda') \cdot \underline{b}_x(\lambda')^T \cdot d\lambda' \right)^{-1} \cdot \underline{b}_x(\lambda) \cdot \left(\int_{\lambda'=0}^{\infty} \underline{s}(\lambda') \cdot \underline{s}(\lambda')^T \cdot d\lambda' \right)^{-1} \cdot \underline{s}(\lambda) \cdot \underline{m}_s \cdot d\lambda$$

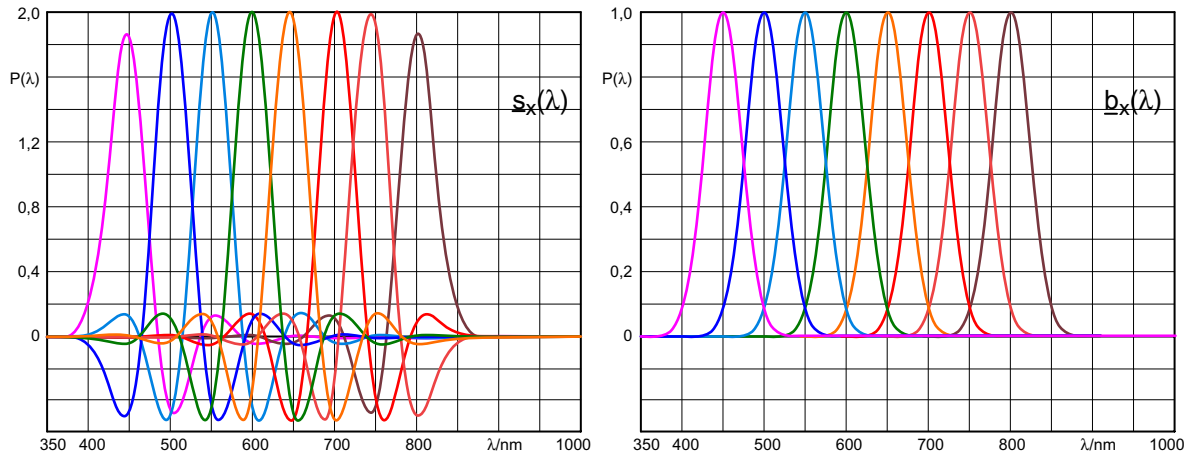


Bild III.10 Spektralwertfunktionen und Primärfarben eines hypothetischen 8-Kanal-System

$$\underline{m}_x = \left(\int_{\lambda'=0}^{\infty} \underline{b}_x(\lambda') \cdot \underline{b}_x(\lambda')^T \cdot d\lambda' \right)^{-1} \cdot \int_{\lambda=0}^{\infty} \underline{b}_x(\lambda) \cdot \underline{s}(\lambda) \cdot d\lambda \cdot \left(\int_{\lambda'=0}^{\infty} \underline{s}(\lambda') \cdot \underline{s}(\lambda')^T \cdot d\lambda' \right)^{-1} \cdot \underline{m}_s$$

$$\underline{m}_x = \underline{\mathfrak{z}} \cdot \underline{m}_s \quad (\text{III.22})$$

Die Ausgaben des Sensors \underline{m}_s werden somit in äquivalente Koordinaten \underline{m}_x des hypothetischen Farbraumes $B_x: (\underline{s}_x(\lambda), \underline{b}_x(\lambda))$ umgerechnet, der bei geeigneter Wahl der Primärfarben eine praktikable Approximation des Umgebungslichtes mit $\phi'(\lambda) = \underline{b}_x(\lambda)^T \cdot \underline{m}_x$ erlaubt.

Bild III.11 zeigt die Rekonstruktion des Spektrums eines Illuminant A aus den Sensorausgaben des 8-Kanal-Sensors nach Bild III.8. Zum Vergleich ist die Approximation dargestellt, die sich direkt aus den Primärfarben des Sensors ergibt und die Approximation unter Verwendung des hypothetischen 8-Kanal-Farbsystems nach Bild III.10. Zu erkennen ist, daß die Näherung mit den hypothetischen Primärfarben einen glatteren Verlauf besitzt und keinen Nulldurchgang aufweist. Da die Primärfarben jedoch den Spektralbereich ab 850 nm nicht abdecken, kann in diesem Bereich keine Approximation erfolgen. Die in beide Fällen erkennbar schlechtere Approximation für Wellenlängen oberhalb 700 nm ist der Sensitivität des Sensors geschuldet, der für diesen Bereich nur zwei verschiedene Empfindlichkeiten aufweist und damit keine ausreichend genaue Rekonstruktion ermöglicht.

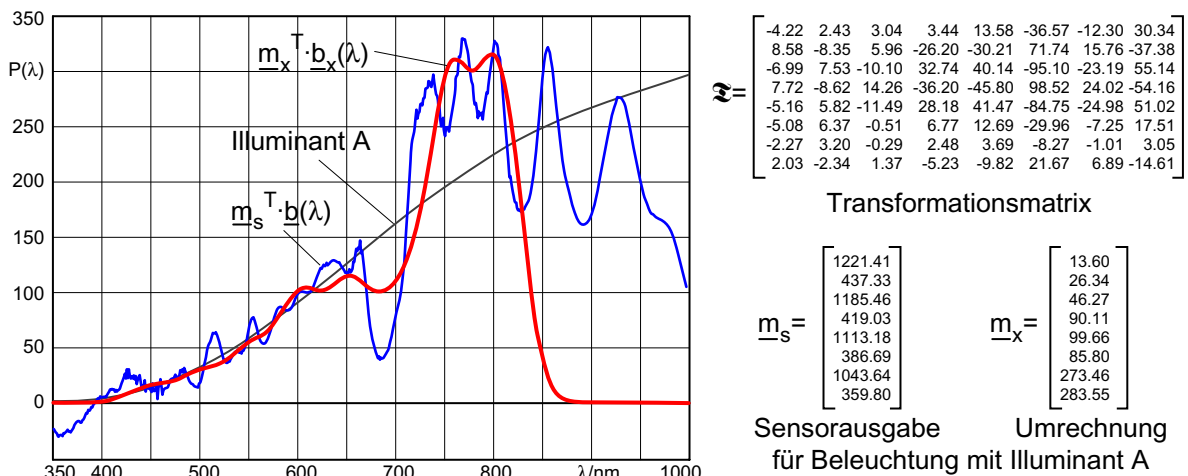


Bild III.11 Approximation des Spektrums von Illuminant A

Mit der Festlegung von geeigneten Rekonstruktionsfunktionen $\underline{b}_x(\lambda)$ kann somit eine hinreichende Approximation des Umgebungslichtes $u'(\lambda)$ aus den Sensorausgaben erfolgen. Diese wird anschließend zur Anpassung der Transformationsparameter verwendet, mit denen die Sensorausgaben in die beleuchtungskompensierten Farbvalenzen umgerechnet werden. Alternativ kann die Näherung der spektralen Leistungsdichte verwendet werden, um eine geeignete Korrekturmatrix aus einer Reihe vorberechneter Matrizen für verschiedene typische Umgebungslichter auszuwählen. Diese Methode der Umgebungslichtkompensation läßt sich dabei direkt in die Farbkorrektur mittels spektraler Rekonstruktion integrieren (Bild III.12). Als analytische Vorgaben sind die Sensorfunktion $\underline{s}(\lambda)$ und die spektrale Empfindlichkeit des angestrebten idealen Sensors $\underline{a}(\lambda)$, das Spektrum des gewünschten Umgebungslichtes, sowie die Auswahl eines geeigneten weißen Bildpunktes aus den gemessenen Sensordaten notwendig.

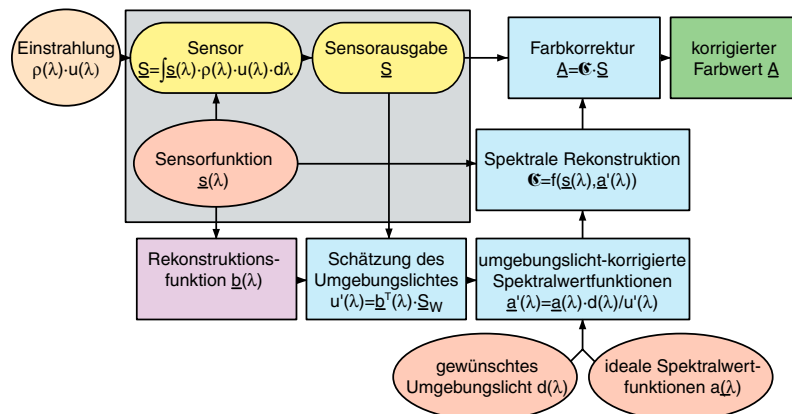


Bild III.12 Blockschaltbild für automatischen Weißabgleich mittels Spektraler Rekonstruktion

5.4 Schlußfolgerungen

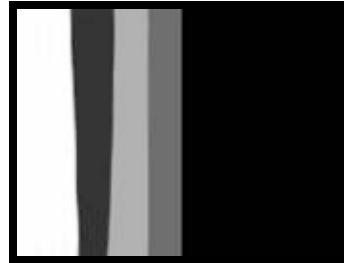
Wie bereits in Abschnitt 2.1 am Beispiel der Vektorraum-Approximation illustriert, ist es nicht möglich, das auf den Sensor wirkende Spektrum aus einem dreidimensionalen menschlichen Farbraum heraus hinreichend genau zu bestimmen. Die Genauigkeit der Approximation aus einem höherdimensionalen Sensorraum ist zwar ebenfalls begrenzt, es ergibt sich jedoch aufgrund dieser höheren Dimension eine bessere Approximation von $u(\lambda)$ und folglich auch eine bessere Kompensation des Umgebungslichtes in der Anwendung.

Die Rekonstruktion des auf den Sensor einwirkenden Lichtspektrums kann zudem benutzt werden, um auch die Farbwiedergabe zu verbessern. Im vorangegangenen Kapitel diente die Approximation des auf den Sensor wirkenden Spektrums zur Kompensation des Umgebungslichtes. Der dazu verwendete höherdimensionale Wiedergaberaum kann jedoch auch verwendet werden, um Lichtspektren nicht nur metamer, sondern auch isomer zu reproduzieren. Damit wird identische Farbwahrnehmung auch bei individuellen Unterschieden in den Sehpigmenten der Betrachter sichergestellt. Erste Anwendungen, die in ihrer Konzeption jedoch eine andere Zielsetzung bei der Wiedergabe haben und den Sensor unbeachtet lassen, wurden bereits in Kapitel II, Abschnitt 3.3 gezeigt.

Nachfolgend sind Beispiele für Farbkorrekturen dargestellt, die durch Systemsimulation verschiedener Sensorkonfigurationen ermittelt wurden. Dabei wurde neben dem Farbbild eine infrarote Störung eingefügt, um die Abhängigkeiten der unterschiedlichen Sensoren in diesem Bereich zu verdeutlichen.



Originalbild



Infrarot-Störung



herkömmlicher 3-Kanal-Sensor
(nach Kapitel III, Abschnitt 4.1)



4-Kanal-Sensor mit Mosaik
aus Rot-Grün-Blau-Weiss
(nach Kapitel III, Abschnitt 4.2)



4-Kanal-Sensor mit Mosaik
aus Rot-Grün-Blau-Teal
und teilweiser Infrarotfilterung
(nach Kapitel III, Abschnitt 4.2)



4-Kanal-Sensor
mit Doppeldiode und
Zweifarbmosaik
(nach Kapitel III, Abschnitt 4.3)



6-Kanal-Sensor
mit Tripeldiode und
Zweifarbmosaik
(nach Kapitel III, Abschnitt 4.4)



8-Kanal-Sensor
mit Doppeldiode und
Vierfarbmosaik
(nach Kapitel III, Abschnitt 4.5)



Idealer CIE-Sensor
bei Illuminant A als
Umgebungslicht



Umgebungslicht-kompensierter
CIE-Sensor bei Illuminant A
als Umgebungslicht
(nach Kapitel III, Abschnitt 5.3)



Kompensation des
CIE-Sensors mit der
Schätzung des Umgebungslichts
(nach Kapitel III, Abschnitt 5.3)

Bild III.13 Ergebnisse der Systemsimulationen verschiedener Sensorkonfigurationen

Kapitel IV

Licht- und Farbsensoren

白黒はつきりさせる。

1 Bildsensoren

Als Bildsensor soll ein System zur ortsauflösenden Messung der Strahlungsleistung von sichtbarem Licht verstanden werden, das der Aufnahme von Bildern dient. Bildsensoren auf Halbleiterbasis bestehen üblicherweise aus einer matrixförmigen Anordnung von separaten Photodetektoren, die allgemein als Pixel¹⁵ bezeichnet werden. Über eine Optik wird ein reelles Bild auf den Bildsensor abgebildet, das von den Photodetektoren in elektrische Signale umgewandelt wird, die proportional zur empfangenen Strahlungsleistung sind. Zum Auslesen des erzeugten Bildes werden die Pixel der Matrix seriell, parallel oder in beliebiger Reihenfolge (*Random Access*) angesteuert. Das aufgenommene Bild wird im allgemeinen in digitale Form überführt und als Matrix gespeichert, deren Elemente den Grau- oder Farbwert eines gemessenen Bildpunktes repräsentieren. Auf einem Ausgabegerät (Bildschirm, Drucker u.ä.) kann anschließend die Reproduktion des aufgenommenen Bildes erfolgen. Dazwischen ist in der Regel eine Bildverarbeitung notwendig, um Fehler bei Aufnahme oder Wiedergabe zu beseitigen oder derart zu vermindern, daß sie nicht mehr wahrgenommen werden. Durch weiterführende Verarbeitungen können bestimmte Eigenschaften oder Strukturen in den Bildern hervorgehoben werden. Beispiele dafür sind Kontraständerung, Kantenverstärkung, Kantendetektion und Segmentierung [Stein93].

Halbleiter-Bildsensoren spielen in den Gebieten Prozeßautomatisierung, Medizintechnik, Drucktechnik und Unterhaltungselektronik seit den 80er Jahren des 20. Jahrhunderts eine wachsende Rolle. Dabei sind die Anforderungen an die Sensoren stetig gestiegen und werden auch in Zukunft steigen. So wird mit der Einführung von HDTV (*High Definition Television*) eine größere Pixelzahl bei gleichzeitig erhöhter Ausleserate erforderlich.

1.1 Kenngrößen von Bildsensoren

Die Charakterisierung von Bildsensoren mit Kenngrößen ermöglicht einen direkten Vergleich von unterschiedlichen Technologien. Für die Anwendung von Bildsensoren ist dabei besonders die Auflösung interessant, die durch die Anzahl der Pixel definiert ist. Ebenso sind Kriterien wie Genauigkeit und Geschwindigkeit für den praktischen Einsatz relevant. Als Empfindlichkeit eines Bildsensors ist bei einem integrierenden Meßverfahren das Verhältnis des Ausgangssignals (üblicherweise als Spannungshub in [V]) zur Strahlungsenergiedichte (in $\text{J}\cdot\text{cm}^{-2}$) definiert. Die Empfindlichkeit hängt dabei im wesentlichen vom Quantenwirkungsgrad η , den möglichen Verlusten und der Verstärkung der Ausleseschaltung ab. Je nach Anwendungsgebiet des Sensors kann es erforderlich sein, Strahlungen in verschiedenen Wellenlängenbereichen zu detektieren. Die Darstellung der Empfindlichkeit bezogen auf die Wellenlänge des eingestrahnten Lichts wird dabei allgemein als spektrale Empfindlichkeit bezeichnet. Neben der Empfindlichkeit stellt die Linearität eine wichtige Eigenschaft dar.

Der Dynamikbereich ist durch das Verhältnis von maximal möglichem zu minimal auswertbarem Signal gegeben. Das Signal kann in Einheiten der Beleuchtungsstärke, der photogenerierten Ladung oder der ausgelesenen Spannung angegeben werden. Wird die Beleuchtungsstärke verwendet, so ergibt sich die optische Dynamik als Quotient aus der Beleuchtungsstärke bei Sättigung zu rauschäquivalenter Beleuchtung. Letztere läßt sich durch Rückrechnung des Ausgangsrauschens im realen Sensor auf eine äquivalente Beleuchtungsstärke an einem idealen

15. Pixel ist ein Kunstwort aus der Abkürzung der englischen Worte *Picture* (*Pics*, *Pix*) und *Element*.

Sensor bestimmen. Wird die ausgelesene Spannung betrachtet (Ausgangsdynamik), so ist das maximale Signal nach oben durch die verwendete Betriebsspannung begrenzt. Das minimale Signal ergibt sich durch den Effektivwert des Rauschens, die Ausgangsdynamik entspricht damit dem Signal-Rausch-Abstand. Besteht ein linearer Zusammenhang zwischen Beleuchtungsstärke und Signalspannung, so sind optische Dynamik und Ausgangsdynamik identisch.

Der Kontrast ist als Verhältnis von maximalem zu minimalem Signal definiert und damit bei globaler Betrachtung mit dem Dynamikbereich identisch. Als Lokalkontrast ergibt sich der Dynamikbereich zwischen zwei benachbarten Pixeln, der durch lokale Koppeleffekte geringer sein kann als zwischen weit entfernten Pixeln. Je größer der Lokalkontrast, um so schärfer können bei gleichbleibender Auflösung Kanten und Konturen aufgenommen werden.

Die maximale Anzahl der Pixel eines Bildsensors und somit die realisierbare Auflösung ergibt sich durch den Quotienten aus der maximalen Sensorfläche und der minimalen Pixelfläche. Die minimale Pixelgröße ist im wesentlichen durch die Entwurfsregeln der verwendeten Technologie gegeben. Andererseits wird die Minimalgröße auch vom Auflösungsvermögen der Optik bestimmt. Die Sensorfläche ist prinzipiell nur durch die Wafergröße, jedoch zusätzlich meist durch die deutlich geringere Retikelgröße der verwendeten Technologie begrenzt. Eine Erhöhung der Gesamtfläche des Sensors hat auch Auswirkungen auf die Ausbeute, die exponentiell mit steigender Chipgröße abnimmt. Der Flächenfüllfaktor gibt an, welcher Anteil an der Gesamtfläche eines Pixels photoelektrisch aktiv ist. Je nach Aufbau sind verschiedene Bereiche im Pixel abgeschattet und tragen folglich nicht zur Photogeneration bei.

Das transiente Verhalten eines Bildsensors lässt sich je nach Anwendungsgebiet durch mehrere Parameter charakterisieren. Die maximale Pixelrate beschreibt, wie viele Pixel pro Zeiteinheit ausgelesen werden können. Sie ist im wesentlichen durch die Geschwindigkeit des Ausleseverfahrens bestimmt. Die maximale Bildwiederholrate gibt an, wie oft der Sensor pro Zeiteinheit komplett ausgelesen werden kann. Die Bildwiederholrate ergibt sich prinzipiell durch den Quotienten von Pixelrate und Pixelzahl, ist jedoch allgemein durch die zur Sammlung der photogenerierten Ladungsträger notwendige Belichtungszeit begrenzt [Böhm97].

1.2 Farbbildsensoren

Um farbige Bilder aufzunehmen, ist eine Farbmessung an jedem einzelnen Punkt des Bildes notwendig. In der Praxis hat sich dabei das Dreibereichsverfahren zur Farberfassung durchgesetzt (siehe auch Kapitel III, Abschnitt 1.2). Um mit einem Helligkeitssensor eine Farbkamera zu realisieren, ist ein optisches System erforderlich, mit dem das einfallende Licht in drei spektrale Bereiche für z.B. Rot, Grün und Blau zerlegt wird, bevor es auf den Bildsensor trifft.

Eine Realisierungsmöglichkeit besteht beispielsweise in einem Filterradd mit verschiedenen Farbfiltern, das sich vor dem Bildsensor dreht. Die Aufnahme der Bilder für die verschiedenen Farben erfolgt dabei zeitversetzt. Dadurch ergeben sich jedoch Nachteile bei der Aufnahme bewegter Bilder. Zur Synchronisierung ist eine aufwendige mechanische Kopplung von Bildsensor und Filterraddmechanismus notwendig, die sich nicht für kompakte Kamerasysteme eignet. Bei Anwendungen mit geringem zeitlichem Auflösungsbedarf, wie bei der Aufnahme von astronomischen Bildern, werden Filterräder jedoch wegen ihrer Flexibilität häufig eingesetzt. Bei der sogenannten Drei-Chip-Kamera werden drei getrennte Bildsensoren verwendet, die zur Detektion je eines spektralen Teilbereiches zuständig sind. Das Licht wird zu diesem Zweck

entweder mit einem Prisma in seine spektralen Bestandteile zerlegt oder fällt durch einen Strahlteiler und drei Farbfilter auf die Bildsensoren. Diese Lösung ist aufwendig, teuer und verlangt neben drei Sensorschaltkreisen eine komplizierte optische Strahlführung sowie die exakte Justierung der Sensoren untereinander. Die Genauigkeit muß dabei im Bereich der Pixelgröße, also bei wenigen μm liegen. Für kostengünstige Produkte ist dieser Ansatz daher ungeeignet. Für die Massenproduktion sind nur Ein-Chip-Kameras verwendbar, bei denen die Farbaufteilung durch Mosaikfilter geschieht, die direkt auf den Bildsensor aufgebracht werden. An einen spezifischen Pixel wird dabei nur ein Farbanteil gemessen, während die anderen Anteile aus den umgebenden Pixeln interpoliert werden müssen. Diese Variante besitzt zwar ein verringertes Auflösungsvermögen und interpolationsbedingte Bildfehler, kommt jedoch zusätzliche Mechanik und Justage aus und stellt die am weitesten verbreitete Bauform für Farbbildsensoren dar.

1.2.1 Farbfiltermosaik

Die bei Ein-Chip-Kameras eingesetzten Mosaikfilter können verschiedenste Formen besitzen. Eine häufig verwendete Filteranordnung ist dabei das Bayer-Mosaik [Bay76], das aus einer Anordnung von je zwei grün-, einem rot- und einem blauempfindlichen Pixel besteht. Ein so aufgebauter Farbbildsensor besitzt jedoch im Vergleich zu einem Schwarz-Weiß-Sensor bei gleichen Abmessungen eine geringere Auflösung, da zur Bestimmung der gesamten Farbinformation mehrere Pixel örtlich zusammengefaßt werden müssen. Bild IV.1 zeigt in der Praxis verwendete Filteranordnungen. Neben Filtern auf RGB-Basis existieren auch Filter in den Komplementärfarben Magenta, Cyan und Gelb. In jüngster Zeit wird auch an Vierfarbsystemen zur Farbaufnahme gearbeitet [HDBT97], [Sony03].

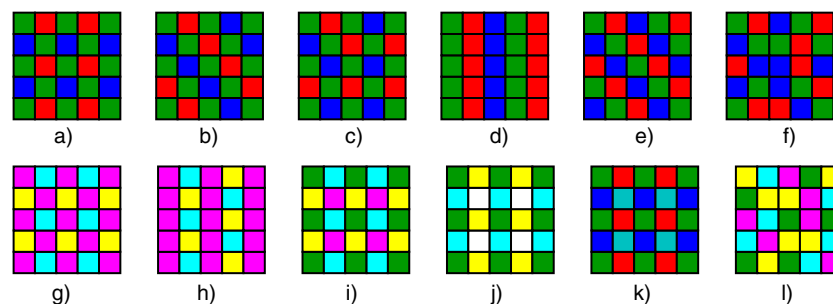


Bild IV.1 Verschiedene Farbfilteranordnungen

a) Bayer-Mosaik, b) Diagonal-Bayer, c) Pseudo-Random-Bayer, d) Vertikal-Linienfilter, e) Diagonal-Filter, f) Pseudo-Random-Filter, g) Komplementär-Mosaik, h) komplementäres Vertikalfilter, i) komplementäres Vierfarbfilter (mit Grün), j) komplementäres Vierfarbfilter (mit Weiß), k) Sony-Vierfarbfilter, l) Pseudo-Random-Vierfarbfilter

Ein allgemeiner Nachteil der Mosaikfilter ist jedoch, daß sie Störungen im Bild bewirken, die abhängig vom Bildinhalt sind. Dazu existiert eine Reihe von Korrekturansätzen, auf die in Abschnitt 6.3 näher eingegangen wird. Weiterhin ist die Empfindlichkeit des Sensors durch die Farbfilterung wesentlich geringer. Zum einen muß bei Farbbildsensoren die Infrarotstrahlung herausgefiltert werden (was bei Schwarz-Weiß-Sensor prinzipiell auch nötig wäre, in der Praxis jedoch nur selten vorgenommen wird) und zum anderen gelangt nur ein Teil des sichtbaren Spektrums zu den Pixeln. Des weiteren liegt die maximale Transmission von praktischen Farbfiltern zwischen 50 und 80%, womit die einfallende Lichtmenge zusätzlich reduziert wird. Farbbildsensoren stellen daher besondere hohe Anforderungen an den Signal-Rausch-Abstand von Sensor und Ausleseschaltung.

1.2.2 Multispektrale Farbbildaufnahme

Die Verwendung von spektralen Farbmeßverfahren zur Bildaufnahme ist in der Praxis auf wenige Spezialfälle begrenzt und wird im allgemeinen nicht zur Aufnahme von farbrichtigen Bildern verwendet. Unter dem Begriff „Multispektrale Farbbildaufnahme“ firmieren dabei sowohl die Spektralanalyse einzelner Pixel als auch die Aufnahme von Bildern mit Filtern verschiedener Wellenlängenbereiche, die z.T. nicht mit den Sensitivitäten des menschlichen Auges übereinstimmen, z.B. Infrarotaufnahmen. Bei einer vollständigen Spektralanalyse kann zudem in der Regel kein komplettes Bild direkt eingelesen werden. Die Systeme verwenden in der Praxis einen Linienscanner, der eine einzelne Zeile der abzubildenden Szene über ein Prisma in seine Spektralanteile zerlegt. Infolge der Dispersion werden die verschiedenen Wellenlängenanteile unterschiedlich stark gebrochen und ergeben mit der abgetasteten Zeile eine zweidimensionale Darstellung. Dieses kann optisch nach Wellenlängenanteilen zusammengefaßt oder direkt auf eine Sensormatrix abbildet werden und erlaubt damit eine spektrale Analyse des einfallenden Lichtes. Ein solches System wird beispielsweise bei Luftaufnahmen eingesetzt. Die zeilenweise Abtastung der Szene erfolgt hierbei durch die Bewegung des Flugzeuges über das aufzunehmende Gelände. In [CSS03] ist Aufnahme von Spektralbildern durch Zerlegung in 224 Kanäle in 10nm-Schritten im sichtbaren und nahen Infrarotbereich beschrieben. Kameras mit unterschiedlichen Filtersätzen werden auch bei astronomischen Beobachtungen und in der Raumfahrt verwendet. Die Filter im Bereich von Ultraviolett bis Infrarot dienen beispielsweise dazu, mit Satelliten Wolken zu erfassen, die Planetenoberfläche trotz Bewölkung bildlich aufzunehmen oder Objekte anhand ihrer spezifischen spektralen Reflektanz zu identifizieren. Ein anders Beispiel für die Anwendung von Filtern sind Multispektral-Infrarot-Thermometer. Auf Basis der Erkenntnisse zur Schwarzkörperstrahlung (siehe auch Kapitel II, Abschnitt 3.4) sind diese in der Lage, Temperaturen sehr genau und berührungslos zu erfassen.

2 Optische Filter für Farbbildsensoren

Die spektrale Empfindlichkeit eines Bildsensors auf Halbleiterbasis wird im wesentlichen von der verwendeten Technologie bestimmt. Für eine korrekte Farbbildaufnahme ist es jedoch essentiell, daß die Luther-Bedingung erfüllt ist (siehe auch Kapitel III, Abschnitt 1). Dies kann durch geeignete optische Filter erfolgen, die sich als Absorptionsfilter oder Interferenzfilter realisieren lassen. In [WS82] sind verschiedene Methoden vorgestellt, deren technische Umsetzung im Schaltkreisentwurf jedoch schwierig ist. Eine Realisierung der Normspektralwertfunktionen als Filtertransmission ist erst vor kurzem mit Interferenzfiltern gelungen [Maz04], wobei jedoch noch keine Produkte existieren.

2.1 Physikalische Grundlagen

Ein optischer Filter ist ein System, das die spektrale Leistungsdichte $\varphi(\lambda)$ eines Lichtstrahls verändert. Liegen die gefilterten Wellenlängen im sichtbaren Spektralbereich, spricht man im allgemeinen von einem Farbfilter. Die beschreibende Größe des Filters ist dabei dessen Transmission. Läuft ein Lichtstrahl mit der spektralen Strahlungsleistung Φ_0^λ durch einen mehrschichtigen Aufbau gemäß Bild IV.2, wird dieser infolge der Wechselwirkung mit den Materialien in einen reflektierten, einen transmittierten und einen absorbierten Anteil aufgespalten. Die Summe von Reflexionsgrad $\rho(\lambda)$, Transmissionsgrad $\tau(\lambda)$ und Absorptionsgrad

$\alpha(\lambda)$ ist dabei aus Gründen der Energieerhaltung gleich Eins:

$$\rho(\lambda) + \tau(\lambda) + \alpha(\lambda) = 1 \quad (\text{IV.1})$$

$$\text{mit:} \quad \tau(\lambda) = \frac{\Phi_{\text{T}}^{\lambda}}{\Phi_0^{\lambda}} \quad \rho(\lambda) = \frac{\Phi_{\text{R}}^{\lambda}}{\Phi_0^{\lambda}} \quad \alpha(\lambda) = \frac{\Phi_{\text{A}}^{\lambda}}{\Phi_0^{\lambda}}$$

Die Beeinflussung der Transmission eines Lichtstrahles durch ein Medium kann nach Gleichung (IV.1) durch die Absorption oder die Reflexion geschehen und führt zu den zwei möglichen Realisierungen optischer Filtern, den Interferenz- und den Absorptionsfiltern.

Das Lambertsche Absorptionsgesetz beschreibt die Schwächung einer spektralen Strahlungsleistung in einem Medium mit dem Absorptionskoeffizienten $\beta(\lambda)$ nach der Wegstrecke x :

$$\Phi^{\lambda}(x) = \Phi^{\lambda}(0) \cdot e^{-\beta(\lambda) \cdot x} \quad (\text{IV.2})$$

Die Reflexion ist ein Grenzflächeneffekt, der auf unterschiedliche Brechungsindizes n_0 und n_1 der aneinandergrenzenden Materialien zurückzuführen ist. Für die Reflexion an einer planparallelen Platte ergibt sich bei Vernachlässigung der Absorption (nach Fresnel):

$$\rho = \left(\frac{n_0 - n_1}{n_0 + n_1} \right)^2 \quad \tau = (1 - \rho)^2 \quad (\text{IV.3})$$

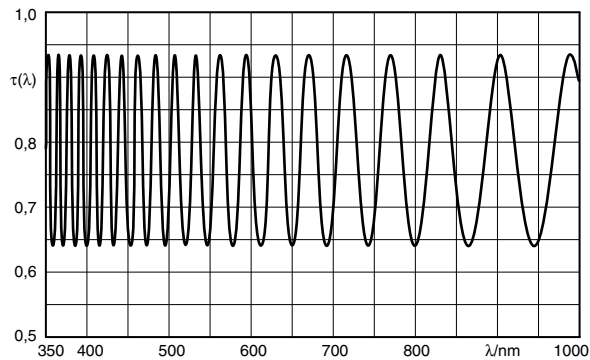
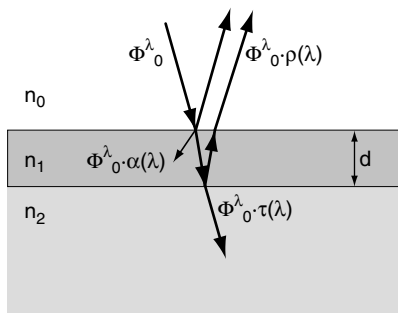


Bild IV.2 Interferenz an einer dünnen Schicht
Modell (links) und Wellenlängenabhängigkeit der Transmission für eine 4 µm dicke SiO₂-Schicht auf Si-Substrat nach Gleichung (IV.4) (rechts)

2.2 Interferenzfilter

Interferenzfilter beruhen auf Interferenzen zwischen einfach oder mehrfach reflektierten Lichtstrahlen. Bei Vernachlässigung der Absorption ergibt sich die Transmission an einem zweischichtigen Aufbau nach Bild IV.2 zu:

$$\tau(\lambda) = \frac{(1 - s_1^2) \cdot (1 - s_2^2)}{1 + s_1^2 \cdot s_2^2 + 2 \cdot s_1 \cdot s_2 \cdot \cos(\varphi)} \quad (\text{IV.4})$$

$$\text{mit:} \quad s_1 = \frac{n_0 - n_1}{n_0 + n_1} \quad s_2 = \frac{n_1 - n_2}{n_1 + n_2} \quad \varphi = 2\pi \cdot n_1 \cdot \frac{d}{\lambda}$$

Die Interferenz der an den Grenzflächen reflektierten Strahlen verursacht somit eine starke Wellenlängenabhängigkeit der Transmission. Interferenzfilter stellen die praktische Anwendung dieses Effektes dar und bestehen aus verschiedenen Schichten mit Dicken im Wellenlängenbereich. Die Schichten der Interferenzfilter können dabei aus transparenten Materialien bestehen, wodurch der Einfluß der Absorption auf die Transmissionsmaxima vernachlässigt werden kann.

Die sogenannten Fabry-Perot-Filter sind schmalbandige Interferenzfilter, die aus nur zwei hochreflektierenden Schichten mit einer transparenten Zwischenschicht bestehen. Die Spiegelschichten können dabei aus dünnen Metallschichten oder aus dielektrischen Vielfachschichten aufgebaut sein. In der Laserspektroskopie dienen solche Filter u.a. dazu, die Laserbandbreite zu verringern [Dem91]. Die Filtercharakteristik ist allgemein von der Kombination von Schichtmaterial und Schichtdicke abhängig. Mit den Methoden der Dünnschichttechnik lassen sich Interferenzfilter auf einem transparenten Substrat oder direkt auf einem Schaltkreis deponieren. Bei der Verwendung von Interferenzfiltern direkt auf dem Schaltkreis ergibt sich jedoch der Nachteil, daß zur Herstellung der Filter Materialien verwendet werden, die eine Kontaminationsquelle für Schaltkreise und Prozeßanlagen darstellen. Daher hat diese Methode noch keine praktische Anwendung gefunden [Schick01].

Bei einem Photosensor sind Interferenzeffekte dennoch von Bedeutung. Da die Oberfläche eines Schaltkreises von verschiedenen herstellungsbedingten Schichten bedeckt ist, treten auch dort Interferenzen auf, die Einfluß auf das optische Verhalten des Sensors haben. Für das Design von Sensoren ist es daher wichtig, die optischen Eigenschaften zu berücksichtigen, die sich aus dem jeweiligen Herstellungsprozeß ergeben (vergl. auch Bild III.5 - Bild III.8).

2.3 Absorptionsfilter

Absorptionsfilter werden aufgrund ihrer preiswerten Herstellung in vielen technischen Bereichen zur Filterung von Licht verwendet. Zur Herstellung der Filter sind hauptsächlich die Einbettung von Farbpigmenten in ein Trägersubstrat und die Beschichtung eines transparenten Substrates mit einer Farbschicht von Bedeutung [Dris78]. Im Falle eines integrierten Photosensors stellt die Silizium-Scheibe mit den entsprechenden Sensoren das Trägermaterial dar, auf das eine farbige Schicht aufgebracht und strukturiert werden kann. Bei der Herstellung von Farbbildsensoren wird fast ausschließlich diese Methode angewendet. Die spektrale Transmission des Filters wird dabei von der Absorption des Farbstoffes bestimmt, die als Materialeigenschaft von dessen chemischer Zusammensetzung abhängig ist. Als Folge können Absorptionsfilter nur in beschränktem Maße angepaßt werden. In der Literatur sind verschiedene Farbfiler für Bildsensoren beschrieben, die jedoch größtenteils von CCD-Herstellern stammen und nicht frei verfügbar sind [Schick01]. Bei Vergleich der publizierten Absorptionsspektren von Filtern verschiedener Hersteller fallen jedoch ähnliche Filtermerkmale auf. Daraus läßt sich schließen, daß die Zahl der in der Praxis verwendeten Materialien für Absorptionsfilter relativ begrenzt ist.

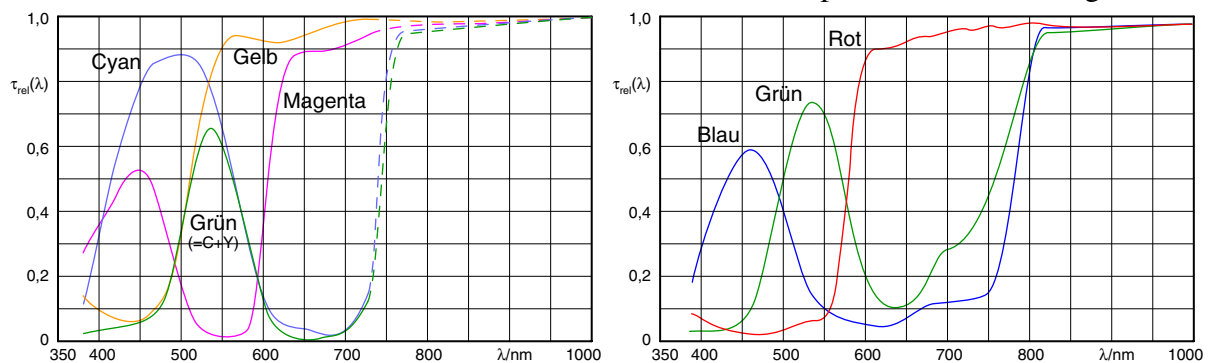


Bild IV.3 Verschiedene Absorptionsfilter

Komplementärfarbfiler aus [Schick01] mit Grün aus der Überlagerung von Cyan und Gelb (links, Infrarotbereich ergänzt) und Primärfarbfiler aus [Mic04] (rechts)

3 Elektrische Verfahren zur Strahlungsmessung

3.1 Der äußere lichtelektrische Effekt

Zur Realisierung von elektronischen Bildsensoren sind elektrische Bauelemente zur Messung von Strahlungsenergie notwendig. Sichtbares Licht ist als elektromagnetische Welle im Tera-Hertz-Bereich mit Wellenlängen zwischen 380 und 700 nm nicht direkt meßbar. Zur Messung der Strahlungsintensität existieren neben photochemischen auch elektrische Verfahren. Entgegen der Vielfalt der möglichen chemischen Reaktionen auf die Einwirkung von Licht, existiert jedoch nur ein physikalischer Effekt, der die Verbindung von Licht und elektrisch meßbaren Größen erlauben: der lichtelektrische Effekt. Der lichtelektrische Effekt, oder Photoeffekt, ist ein physikalischer Vorgang, bei dem Licht seine Energie vollständig an Elektronen übergibt. Beim äußeren Photoeffekt treten die Elektronen nach der Energieaufnahme aus der Oberfläche des beleuchteten Materials. Beispielsweise entlädt sich eine negativ geladene Metallplatte unter Lichteinwirkung. Die ersten Beobachtungen zum Photoeffekt stammen aus dem Jahr 1887 von Heinrich Hertz. Wilhelm Ludwig Franz Hallwachs untersuchte 1888 diesen Effekt genauer und schuf den Grundversuch zum äußeren photoelektrischen Effekt mit der negativ geladene Zinkplatte und der Bestrahlung durch UV-Licht. Die Erklärung des Effektes gelang jedoch zu dieser Zeit nicht, da man bis dahin alle Naturphänomene von den Newtonschen Axiomen und der Maxwell'schen Wellentheorie ableitete. Erst 1905 konnte der Photoeffekt von Albert Einstein mit der Quantentheorie des Lichts erklärt werden. Danach besitzt ein Lichtquant der Frequenz f die Energie $E = h \cdot f$, wobei h das Plancksche Wirkungsquantum darstellt. Beim äußeren Photoeffekt tritt ein Elektron der Masse m_e aus dem Material aus, wenn die vom Photon übertragene Energie größer als die materialabhängige Austrittsarbeit W_A ist. Allgemein gilt:

$$h \cdot f = W_A + \frac{m_e \cdot v^2}{2} \quad (\text{IV.5})$$

mit der Geschwindigkeit v des Elektrons nach dem Austritt aus dem Metall. Mit diesem Modell widersprach Einstein jedoch der klassischen Physik. Erst nachdem Robert Andrews Milikan 1916 die Gültigkeit von Einsteins Aussagen experimentell nachweisen konnte, erhielt Einstein unter anderem für diese Leistung 1921 den Nobelpreis für Physik. Die Entdeckung des lichtelektrischen Effekts stellt somit den Beginn der modernen Quantenphysik dar.

3.2 Vakuumphotozelle und Sekundärelektronenvervielfacher

Anwendungen des äußeren lichtelektrischen Effektes stellen beispielsweise die Vakuumphotozelle und der Sekundärelektronenvervielfacher dar. Eine Vakuumphotozelle besteht aus einer luftleeren Glaskugel, in der sich neben einer Fotokathode eine Anode zur Aufnahme der freiwerdenden Photoelektronen befindet. Als Kathodenmaterial eignen sich Kombinationen verschiedener Alkali- und Erdalkalimetalle, die mit der notwendigen Austrittsarbeit die Obergrenze des detektierbaren Wellenlängenbereiches bestimmen. In die Photozelle einstrahlendes Licht löst Elektronen aus der Kathode. Die nun positiv geladene Kathode stößt die freien Elektronen in Richtung Anode ab, die bei geschlossenem Stromkreis als Photostrom gemessen werden können. Mit einer Bremsspannung zwischen Anode und Katode, die dem Photostrom entgegenwirkt, lassen sich z.B. das Plancksche Wirkungsquantum h , die Austrittsarbeiten für das Kathodenmaterial und die entsprechenden Grenzwellenlängen bestimmen [HMS95]. Ein

Sekundärelektronenvervielfacher (Photomultiplier) dient dazu, schwache Lichtimpulse in elektrische Ladungsimpulse umzuwandeln und zu verstärken (ca. 10^7 -fach). Wie bei einer Vakuumphotozelle werden an der Katode Photoelektronen durch den äußeren Photoeffekt ausgelöst. Die freigesetzten Elektronen werden anschließend durch eine angelegte Hochspannung zur ersten Dynode des Vervielfachersystems beschleunigt und setzen dort weitere Sekundärelektronen frei. Dies wiederholt sich, bis die Elektronenlawine zur Anode gelangt. Der Anodenstrom ist dabei proportional zu der Lichtmenge, die auf der Photokathode auftrifft.

3.3 Halbleiter-Photosensoren

Bei Halbleitern tritt der innere Photoeffekt auf. Dieser beruht auf demselben Prinzip wie der äußere Photoeffekt, die Elektronen treten hier jedoch nicht aus der Oberfläche des Materials aus. Wird ein Halbleiter beleuchtet, so nehmen die Elektronen die Energie der Lichtquanten auf und erhöhen ihr Energieniveau. Ist die aufgenommene Energie größer als die Bandlücke, so wechseln die Elektronen aus dem Valenzband in das Leitungsband und stehen als frei beweglichen Ladungsträger zur Verfügung. Photosensoren auf Halbleiterbasis beruhen auf dem inneren lichtelektrischen Effekt. Der innere Photoeffekt bildet somit die Basis für die moderne Strahlungsmessung, die Photovoltaik und die Digitalkameratechnik. Halbleiterphotosensoren sind gegenüber der Röhrentechnik nicht nur funktionsmäßig und preislich im Vorteil, sondern auch wegen der Möglichkeit zur Integration. Integrierte Photosensoren zeichnen sich gegenüber anderen Systemen durch höhere Empfindlichkeit und eine wesentlich geringere Verlustleistung aus. In digitale Systeme eingebettete Halbleitersensoren bieten außerdem gute Kompatibilität zu Rechensystemen. Dadurch werden die Verarbeitung von Bildsignalen und die Verwendung von konventionellen Speicherbausteinen zur Speicherung von digitalen Bildern möglich.

3.3.1 Optische Generation

Die Bestrahlungsstärke E ist ein Maß für die pro Fläche auftreffende Strahlungsleistung. Durch Bezug auf die Wellenlänge λ der Strahlung erhält man die in $[\text{W} \cdot \text{m}^{-2} \cdot \text{nm}^{-1}]$ gemessene spektrale Bestrahlungsstärke E^λ . Die spektrale Photonenflußdichte Π^λ in $[\text{m}^{-2} \cdot \text{s}^{-1} \cdot \text{nm}^{-1}]$ wird daraus ableitend mit der Anzahl der Photonen pro Fläche, Zeit und Wellenlänge angegeben. Zwischen E^λ und Π^λ besteht folgender Zusammenhang:

$$\Pi^\lambda = \frac{1}{h \cdot f} \cdot E^\lambda = \frac{\lambda}{h \cdot c} \cdot E^\lambda \quad (\text{IV.6})$$

mit dem Planckschen Wirkungsquantum h , der Frequenz f und der Lichtgeschwindigkeit c . Trifft eine elektromagnetische Strahlung auf einen Festkörper, so wird ein Teil der Strahlung an der Oberfläche reflektiert und ein weiterer absorbiert. Der verbliebene Anteil wird transmittiert. Die Absorption des Photonenflusses an einem Ort x im Festkörper ist proportional zur dort vorhandenen Photonenflußdichte Π^λ . Die Funktion der Photonenflußdichte über der Wellenlänge sei mit $\pi(\lambda)$ definiert, die ortsabhängige Photonenflußdichte mit $\pi(\lambda, x)$. Die Absorption im Festkörper wird durch das Lambertsche Gesetz beschrieben. Im eindimensionalen Fall eines zur Oberfläche senkrechten Strahlungseinfalls gilt:

$$\pi(\lambda, x) = \pi(\lambda, 0) \cdot e^{-\beta(\lambda) \cdot x} \quad (\text{IV.7})$$

$\pi(\lambda, 0)$ stellt dabei die Photonenflußdichte an der Oberfläche des Festkörpers abzüglich des re-

flektierten Anteils dar. Die Absorption erfolgt für Photonen unterschiedlicher Wellenlängen unterschiedlich stark. Dies wird durch den materialabhängigen Absorptionskoeffizienten $\beta(\lambda)$ beschrieben (Bild IV.4). Der Abstand $\beta(\lambda)^{-1}$ wird dabei als Eindringtiefe bezeichnet, da für $x = \beta(\lambda)^{-1}$ die Photonenflußdichte der Wellenlänge λ um den Faktor e^{-1} absinkt. Die Eindringtiefe wächst allgemein mit der Wellenlänge des Lichtes. Photonen mit hoher Energie (geringer Wellenlänge) werden oberflächennah absorbiert, während Photonen mit größeren Wellenlängen tiefer in den Halbleiter eindringen und somit über einen größeren räumlichen Bereiche absorbiert werden.

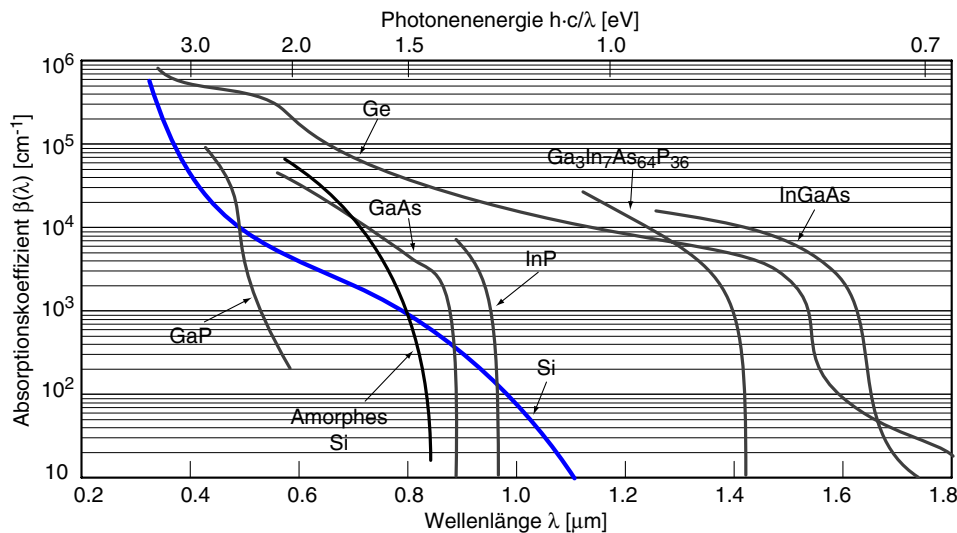


Bild IV.4 Verlauf der Absorptionskoeffizienten über die Wellenlänge für verschiedene Materialien

Das Absorptionsverhalten ist weiterhin stark vom bestrahlten Material abhängig. Metalle absorbieren Photonen innerhalb eines großen Wellenlängenbereiches, da zahlreiche Elektronen im Leitungsband vorhanden sind, die durch die Wechselwirkung mit den Photonen auf nahezu beliebige höhere Energieniveaus angehoben werden und den Festkörper z.T. auch verlassen können. In Halbleitern sind jedoch nur wenige freie Ladungsträger in den Bändern vorhanden. Hier werden Photonen überwiegend dadurch absorbiert, daß durch optische Generation, oder Photogeneration, zusätzliche Elektron-Loch-Paare generiert werden. Das einfallende Photon muß zur Generierung von Ladungsträgern im intrinsischen Halbleiter eine Energie $h \cdot f$ besitzen, die größer als der Bandabstand ΔW des Materials ist, ansonsten transmittiert es fast ungehindert durch das Material. Im dotierten Halbleiter existiert zusätzlich die Möglichkeit, daß ein Elektron die geringere Energiedifferenz zwischen Leitungsbandkante und Donatorniveau überwindet. In allen Fällen ergibt sich eine maximale Wellenlänge, die zur Photogeneration beiträgt.

$$\lambda_{\min} = \frac{h \cdot c}{\Delta W} \quad (\text{IV.8})$$

Jedoch führt nicht jedes absorbierte Photon zu einem Elektron-Loch-Paar. So können Photonen mit zu großer Energie und damit geringer Eindringtiefe direkt an der Oberfläche absorbiert werden, wo sie aufgrund von Oberflächenzuständen, die als Rekombinationszentren wirken, sofort wieder rekombinieren. Aus diesem Grund existiert auch eine untere Grenzwellenlänge der Photogeneration. In Halbleitern werden in geringem Maße auch Photonen mit geringerer Energie als der Bandabstand absorbiert. Diese tragen jedoch nicht zur Photogeneration bei, sondern heben wie in Metallen jeweils eines der wenigen freien Elektronen auf ein höheres Energieniveau.

Beim Rückfall auf ein niedrigeres Niveau wird die Differenzenergie in Form von Gitterschwingungen als Wärmeenergie abgegeben.

Die Zahl der generierten Ladungsträgerpaare bezogen auf die Zahl der absorbierten Photonen wird als spektraler Quantenwirkungsgrad $\eta_q(\lambda)$ bezeichnet, der maximal den Wert 1 erreichen kann. Die spektrale optische Generationsrate $G(\lambda, x)$ ergibt sich durch Differentiation der spektralen Photonenflußdichte nach x unter Berücksichtigung des Quantenwirkungsgrades:

$$G(\lambda, x) = -\eta_q(\lambda) \cdot \frac{\partial}{\partial x} \pi(\lambda, x) = \eta_q(\lambda) \cdot \pi(\lambda, 0) \cdot \beta(\lambda) \cdot e^{-\beta(\lambda) \cdot x} \quad (\text{IV.9})$$

Die Fotogeneration hängt somit von materialabhängigen Koeffizienten, von der Tiefe x im Halbleiter sowie von der Dichte der einfallenden Photonen und deren Wellenlänge ab. Für kristallines Silizium ergibt sich beispielsweise ein Bereich der einfallenden Strahlung von etwa 300 nm bis 1200 nm, in dem eine nennenswerte Photogeneration stattfindet.

Die Anzahl der freien Ladungsträger eines Halbleiters befindet sich jedoch im Gleichgewicht. Mit der Rekombination existiert ein Mechanismus, welcher der Ladungsträgergeneration entgegenwirkt. Die Rekombinationsrate ist dabei proportional zur Anzahl der freien Ladungsträger. In einem bestrahlten Halbleiter rekombinieren die photogenerierten Ladungsträger und geben dabei je nach Rekombinationsmechanismus Strahlung oder Wärmeenergie an das Gitter ab. Eine Messung der Lichtintensität ist somit allein durch die optische Generation nicht möglich. Zur technischen Nutzung der Photogeneration, wie beispielsweise zur photovoltaischen Energieerzeugung oder zur Detektion von Strahlung, müssen die generierten Elektronen-Loch-Paare getrennt werden, um deren Rekombination zu verhindern. Zur Trennung von Ladungsträgern ist ein elektrisches Feld geeignet. Dies kann durch Anlegen einer externen elektrischen Spannung oder durch Erzeugung einer Raumladungszone geschehen.

3.3.2 Photowiderstand

Ein Photowiderstand ist ein mit zwei ohmschen Kontakten versehener homogener dotierter Halbleiter. Das Bauelement ist in jedem Betriebszustand neutral. Es treten keine Bandverbiegung und damit verbundenen Raumladungen auf. Ein Photowiderstand wirkt als ohmscher Widerstand, dessen Leitfähigkeit proportional zur Zahl der freien Ladungsträger ist.

Wird im unbeleuchteten Fall eine Spannung angelegt, so fließt ein temperaturabhängiger Dunkelstrom I_0 . Im beleuchteten Fall bewirkt die angelegte Spannung eine Trennung der photogenerierten Ladungsträger und ein Photostrom trägt zusätzlich zum Dunkelstrom zum Stromfluß bei. Für einen n-Halbleiter ergibt sich die so erzeugte spektrale Photostromdichte $j_{ph}(\lambda)$ zu:

$$j_{ph}(\lambda) = e \cdot \eta_q(\lambda) \cdot \pi(\lambda, 0) \cdot (1 - e^{-\beta(\lambda) \cdot d}) \cdot \frac{\tau_n}{\tau_t} \quad (\text{IV.10})$$

Dabei ist τ_n die Lebensdauer der Majoritätsladungsträger und τ_t ihre Transitzeit. Der Quotient τ_n / τ_t wird dabei als Gewinn bezeichnet und gibt die Anzahl der Ladungsträger an, die je Generationsvorgang zum Stromfluß beitragen. Ist die Lebensdauer größer als die Transitzeit, was für realistische Dotierungskonzentrationen und Geometrien meist der Fall ist, wird der Gewinn größer als 1, in kristallinem Silizium ist $\tau_n / \tau_t \approx 10^3$ erreichbar.

Die Anzahl der photogenerierten Ladungsträger kann im Photowiderstand somit die Anzahl der Photonen im selben Zeitraum um ein Vielfaches übersteigen, da ein einzelnes photogeneriertes Elektron, das die Anode erreicht, ohne vorher zu rekombinieren, aufgrund der Ladungsneutralität

lität des Photowiderstands die Injektion eines Elektrons aus der Kathode bewirkt. Dieser Prozeß setzt sich auch ohne erneute Photogeneration so lange fort, bis das Elektron rekombiniert. Der Gewinn gibt somit die Anzahl der Elektronendurchläufe pro Generationsvorgang an. Die Lebensdauer und die Transitzeit stellen jedoch Mittelwerte dar und die Zahl der effektiv generierten Ladungsträger ist nur statistisch zu ermitteln. Aufgrund der Zeitabhängigkeit, der Nichtlinearität sowie der Abhängigkeit von anderen Parametern wie beispielsweise der Temperatur, eignet sich der Photowiderstand nicht für genaue Messungen der Strahlungsintensität. Photowiderstände bieten sich jedoch als Photodetektoren für Anwendungen an, die Licht mehr oder weniger diskret verwenden, wie beispielsweise Lichtschranken, Lichtrelais, Dämmerungsschalter, Zähleinrichtungen, Lichtüberwachung o.ä.

3.3.3 Photodiode

Eine Photodiode ist eine mehrschichtige Halbleiterstruktur, die eine oder mehrere Potentialbarrieren aufweist. Durch das elektrische Feld in den entsprechenden Raumladungszonen werden die durch Strahlungseinwirkung generierten Ladungsträger separiert und driften in entgegengesetzte Gebiete. Dieser Photostrom kann anschaulich als Verschiebung der U-I-Kennlinie der Diode dargestellt werden (Bild IV.5). Dabei wird beispielsweise in der Photovoltaik der Betrieb der Diode im 4. Quadranten zur Umwandlung von Strahlung in elektrische Energie ausgenutzt. Für die Strahlungsmessung interessanter ist der Betrieb der Photodiode in Sperrrichtung im 3. Quadranten. Im unbeleuchteten Fall fließt bei Anlegen einer Sperrspannung der Sperrstrom (Dunkelstrom) I_0 , zu dem sich bei Beleuchtung der Photostrom I_{ph} addiert.

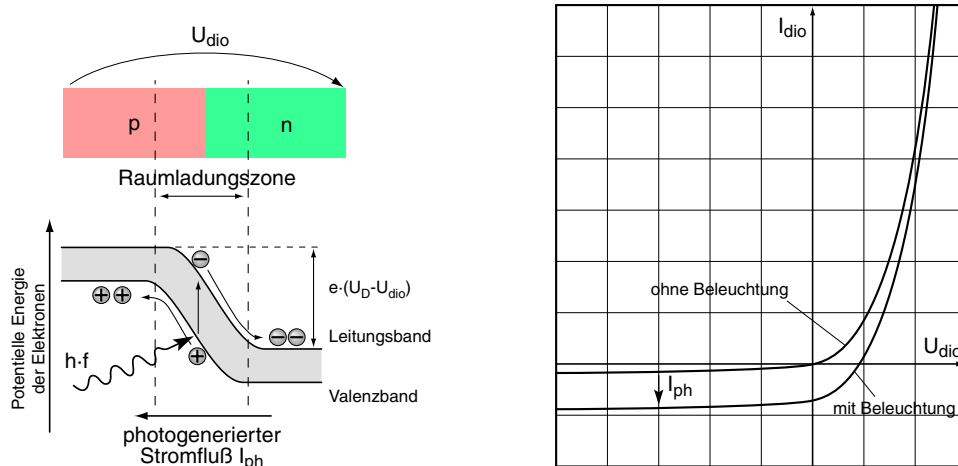


Bild IV.5 Halbleitervorgang bei Bestrahlung und U-I-Kennlinie einer Photodiode

Zur Bestimmung der spektralen Photostromdichte $j_{ph}(\lambda)$ muß das Halbleitergleichungssystem unter Berücksichtigung der Photogeneration gelöst werden. Eine gebräuchliche Näherung ergibt sich unter der Annahme, daß alle photogenerierten Ladungsträger zum Stromfluß beitragen. Die spektrale Photostromdichte entspricht dann dem Integral der Generationsrate $G(\lambda, x)$ über die jeweilige Dicke der photoelektrisch aktiven Schicht des Halbleiters:

$$j_{ph}(\lambda) = e \cdot \int_{x_0}^{(x_0 + d)} G(\lambda, x) \cdot dx = e \cdot \eta_q(\lambda) \cdot \pi(\lambda, 0) \cdot (1 - e^{-\beta(\lambda) \cdot d}) \cdot e^{-\beta(\lambda) \cdot x_0} \quad (IV.11)$$

Die gesamte Photostromdichte j_{ph} kann durch Integration über die Wellenlänge berechnet werden. Dazu müssen die Photonenflußdichte $\pi(\lambda, 0)$ der einfallenden Strahlung sowie die materialspezifischen Wellenlängenabhängigkeiten von $\eta_q(\lambda)$ und $\beta(\lambda)$ bekannt sein. Die Dicke d

ist allgemein gegeben durch den Bereich des jeweiligen Bauelements, in dem eine genügend hohe Feldstärke vorhanden ist, um die generierten Ladungsträger zu trennen. Für einen abrupten pn-Übergang berechnet sich diese Sperrschichtdicke d_s nach [Mö92] zu:

$$d_s = \sqrt{\frac{2 \cdot \varepsilon_H \cdot (N_A + N_D) \cdot (U_D - U_{dio})}{e \cdot N_A \cdot N_D}} \quad \text{mit} \quad U_D = U_T \cdot \ln \frac{N_A \cdot N_D}{n_i^2} \quad (\text{IV.12})$$

Aus Gleichung (IV.11) ist ersichtlich, daß der Photostrom mit zunehmender Dicke wächst. Eine große Sperrschichtdicke ist somit vorteilhaft, um die ortsabhängige Generationsrate optimal zu nutzen. Voraussetzung dafür ist jedoch, daß die photogenerierten Ladungsträger aufgefangen werden können, bevor sie rekombinieren. Dazu muß ihre Driftlänge $l_D = \mu \cdot \tau \cdot E$ größer als d sein. Die Driftlänge wird wiederum von der Beweglichkeit der jeweiligen Ladungsträger μ , deren Lebensdauer τ und der elektrischen Feldstärke E bestimmt und stellt als Beschreibung der Driftbewegung der Ladungsträger den mittleren Weg eines Ladungsträgers dar, der in Gegenwart eines elektrischen Feldes bis zur Rekombination zurückgelegt wird.

Weiterhin ist zu erkennen, daß die spektrale Empfindlichkeit der Photodiode von der Tiefe x_0 abhängig ist, in der sich die Raumladungszone befindet. Tiefe und Dicke, sowie damit verbunden auch die spektrale Empfindlichkeit des Bauelements, sind technologisch über das Dotierungsprofil und in Grenzen elektrisch über die Sperrspannung einstellbar. Beispielsweise können Photodioden als Wannen-Substrat-Dioden in großen Tiefen oder als Diffusionsdioden nahe der Substratoberfläche realisiert werden. Auch eine Kombination von verschiedenen Dioden ist möglich. *Dual-Junction*-Dioden besitzen dabei zwei pn-Übergänge, *Triple-Junction*-Photodioden sind mit drei übereinanderliegenden pn-Übergängen ausgestattet. Für derartige Multikanal-Photodioden besteht jedoch die Forderung, daß sie ähnliche absolute Empfindlichkeiten aufweisen und sich im wesentlichen nur in ihrem spektralen Verlauf unterscheiden. Werden die Photodioden in unterschiedlichen Tiefen implantiert, ist aus Gleichung (IV.11) und Gleichung (IV.12) ersichtlich, daß die Dioden zur Einhaltung dieser Forderung ähnliche Sperrschichtdicken und damit ähnliche Dotierungskonzentrationen aufweisen müssen. Die Photondichte nimmt generell mit der Eindringtiefe ab, somit ist auch für Strahlung mit größeren Wellenlängen eine verringerte Empfindlichkeit in großen Substrattiefen zu erwarten. In natürlichen Umgebungen wird dies jedoch meist durch eine größere Intensität von roter und infraroter Strahlung kompensiert. Nachteilig ist jedoch, daß die Dotierungskonzentrationen technologisch bedingt sind und demzufolge nicht im Entwurfsprozeß beeinflusst werden können. Die Kenntnis von Dotierungsprofilen erlaubt allerdings eine generelle Aussage über die Eignung einer bestimmten Technologie zur Realisierung von Multikanal-Photodioden und entsprechenden Sensoren.

Die mit Photodioden erreichbaren Photoströme sind jedoch generell gering. Eine Beleuchtungsstärke von 500 lx, die vom Auge als sehr hell empfunden wird, entspricht ca. $2.05 \cdot 10^6$ Photonen von 555 nm Wellenlänge, die pro Sekunde auf eine Fläche von $1 (\mu\text{m})^2$ eintreffen. Bei einem typischen Absorptionskoeffizienten $\beta(555\text{nm})$ von $5 \cdot 10^3 \text{ cm}^{-1}$ ($= 0.5 \mu\text{m}^{-1}$) für Silizium ergeben sich $1.03 \cdot 10^6$ generierte Elektronen-Loch-Paare pro $(\mu\text{m})^3$ und Sekunde, wenn die Quanteneffizienz mit 100% angenommen wird. Bei Vernachlässigung der Rekombination und den Annahmen, daß alle Elektronen zum Stromfluß beitragen und das photoaktive Volumen eine Ausdehnung von $10 \times 10 \times 1 \mu\text{m}^3$ besitzt, ergibt sich ein photogenerierter Strom von ca. 16 pA.

3.3.4 Photogate

Da die generierten Photoströme sehr klein sind, stellt ihre genaue Messung ein Problem dar. Viele Verfahren zur Strahlungsmessung beruhen daher auf dem Prinzip der Ladungsakkumulation. Dabei werden die Photoelektronen über eine Integrationszeit gesammelt und anschließend die Ladungsmenge ausgewertet. Die Akkumulation erfolgt dabei üblicherweise in einer Kapazität. Bei einem Photogate wird dazu die MOS-Kapazität genutzt, die das Gate eines Feldeffekttransistors gegenüber dem Substrat besitzt. Aufgrund des elektrischen Feldes zwischen Gate und dem Substrat wird dabei der gesamte Raum unterhalb des Gates zum Sammeln der photogenerierten Ladungsträger genutzt. Die Anzahl der aufzusammelnden Ladungsträger entspricht bei Vernachlässigung von Rekombination, Driftlänge und Lebensdauer der Anzahl der absorbierten Photonen und ergibt sich durch Integration der mit der Quanteneffizienz gewichteten Photonenflußdichte über die Wellenlänge und die beleuchtete Fläche. Bei homogener Bestrahlung über einer Fläche A vereinfacht sich die Beziehung zu:

$$N = \int_A \int_{\lambda} \eta_q(\lambda) \cdot \pi(\lambda, 0) \cdot dA \cdot d\lambda = A \cdot \int_{\lambda} \eta_q(\lambda) \cdot \pi(\lambda, 0) \cdot d\lambda \quad (\text{IV.13})$$

Werden die Ladungsträger über eine Integrationszeit Δt akkumuliert, ergibt sich die gesammelte Ladung bei zeitlich konstanter Bestrahlung zu:

$$Q = \int_{t_0}^{t_0 + \Delta t} e \cdot N \cdot dt = e \cdot \Delta t \cdot A \cdot \int_{\lambda} \eta_q(\lambda) \cdot \pi(\lambda, 0) \cdot d\lambda \quad (\text{IV.14})$$

Die unter dem Photogate gesammelte Ladung wird beim Auslesen üblicherweise auf eine zweite Kapazität C übertragen. Der sich ergebende Spannungshub $\Delta U = Q/C$ wird an einen Ausgangsverstärker weitergeleitet. Aufgrund der Trennung von Integrations- und Auslesekapazität ergibt sich die Möglichkeit der Auslese mit korrelierter Doppelabtastung zur Rauschunterdrückung. Da die Photonen jedoch zuerst das Gate durchdringen müssen, bevor Elektron-Loch-Paare erzeugt werden, ist die Empfindlichkeit dieses Aufbaus besonders für niedrige Wellenlängen geringer als bei direkter Bestrahlung. In der Praxis wird das Gate daher sehr dünn realisiert oder die Bestrahlung von der Rückseite eines abgedünnten Wafers vorgenommen.

3.3.5 Integrierender Photodetektor mit Photodiode

Auch Photodioden lassen sich zum Aufbau eines integrierenden Photodetektors verwenden. Die Photoelektronen werden dabei auf der Sperrschichtkapazität der Photodiode akkumuliert. Die Sperrschichtkapazität ist von der Spannung U über der Diode abhängig und beträgt für einen abrupten pn-Übergang über der homogenen Fläche A ohne Berücksichtigung der Randeffekte:

$$C_{sp} = A \cdot \sqrt{\frac{\epsilon_H \cdot e \cdot N_A \cdot N_D}{2 \cdot (N_A + N_D) \cdot U_D}} \cdot \sqrt{\frac{U_D}{U_D - U}} = C_0 \cdot \sqrt{\frac{U_D}{U_D - U}} \quad (\text{IV.15})$$

mit $U_D = U_T \cdot \ln \frac{N_A \cdot N_D}{n_i^2}$

Bei konstantem Photostrom I_{ph} ist der zeitliche Spannungsverlauf über der Diode bestimmt durch die Differentialgleichung:

$$C_{sp}(U) \cdot \frac{dU(t)}{dt} = -I_{ph} \quad (\text{IV.16})$$

und deren Lösung (Bild IV.6):

$$U(t) = U(0) - \frac{(I_{ph} \cdot t)^2}{4 \cdot U_D \cdot C_0^2} + \frac{I_{ph} \cdot t \cdot \sqrt{U_D^2 - U_D \cdot U(0)}}{U_D \cdot C_0} \quad (IV.17)$$

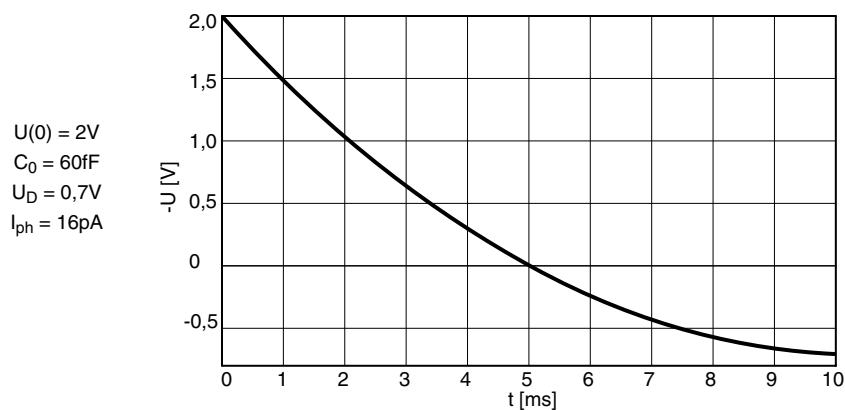


Bild IV.6 Darstellung des zeitlichen Verlaufs der Diodenspannung nach Gleichung (IV.16)

Der Verlauf der Diodenspannung ist für hinreichend kurze Zeiten in Näherung linear und ermöglicht so die Annahme einer linearen Beziehung zwischen der Bestrahlungsstärke und dem meßbaren Spannungswert. Zu beachten ist, daß sowohl Photostrom I_{ph} als auch die Kapazität C_{sp} proportional zur Diodenfläche sind. Die Spannung über der Diode ist somit bei gegebener Beleuchtungsstärke nicht abhängig von Diodenfläche. Dies ist jedoch nur in erster Näherung korrekt. Eine Photodiode besitzt nicht nur eine flächenbezogene Kapazität, sondern auch periphere Kapazitäten. Zudem sind Kapazitäten in externen Schaltungen vorhanden. Mit Verkleinerung der Diodenfläche verringert sich folglich der Photostrom stärker als die Kapazität und damit sinkt auch der erreichbare Spannungshub. Zudem erhöht sich der Einfluß von flächenunabhängigen Rauschtermen und das SNR sinkt. Die Größe der Photodiode bestimmt somit entscheidend den Dynamikbereich des Pixels und sollte daher maximal ausgelegt werden.

4 Pixelschaltungen auf CMOS-Basis

4.1 CCD und CMOS

Zur Nutzung von Halbleiterphotosensoren ist es notwendig, das Signal des photoempfindlichen Elements auszulesen und einer Auswertung zuzuführen. Dafür existieren zwei prinzipiell verschiedene Ansätze, CMOS und CCD¹⁶.

Analoge CMOS-Technik erlaubt, den Photostrom bzw. die akkumulierte Photoladung zu verstärken, transformieren, mit anderen Werten zu verrechnen, zu digitalisieren, etc. [DHBF00]. Die Sensoren können dabei gemeinsam mit den Verarbeitungselementen auf einem Schaltkreis integriert werden und ermöglichen so kostengünstige Sensorsysteme [RKGK99], [Hupp00] (Bild IV.7). Aufgrund von technologischen Mängeln der frühen Fertigungsverfahren war die CMOS-Technik jedoch für die Verarbeitung der geringen Signalpegel der Photosensorik nicht geeignet. Das Problem ist besonders in der herstellungsbedingten Variation der verwendeten Bauelemente begründet. Durch die Schwellspannungsvariation der Transistoren, die Dunkelströme der Photodioden, die zusätzlichen Leckströmen und Parasitärkapazitäten besitzt jede Pixelzelle eine individuelle Charakteristik, die sich in einem sogenannten *Fixed Pattern Noise*

16. *Charge Coupled Device* - Ladungsgekoppeltes Element

(FPN) äußert. Der schlechte Signal-Rausch-Abstand der CMOS-Technik führte daher zu einer verstärkten Konzentration auf die CCD-Technik, die sich in der heutigen Dominanz von CCD auf dem Gebiet von Photosensoren widerspiegelt. Erst mit den technologischen Verbesserungen des Herstellungsprozesses und den Möglichkeiten von algorithmischen Fehlerkorrekturen [Rull02] wurden Photosensoren auf CMOS-Basis wieder interessant.

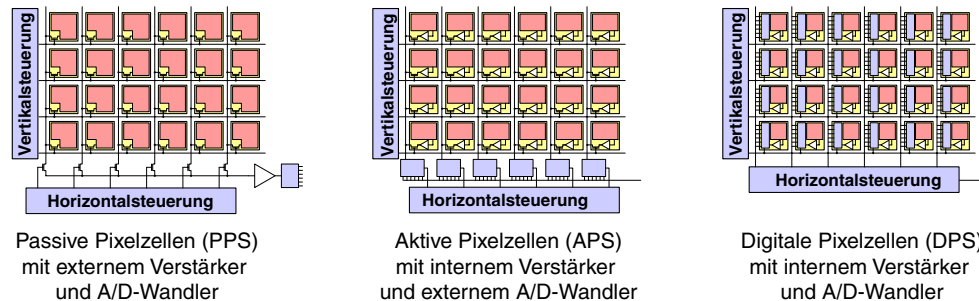


Bild IV.7 Verschiedene Varianten für den Aufbau eines CMOS-Bildsensors

Die CCD-Technik basiert auf den Entwicklungen zu Analogspeichern auf Basis von MOS-Kapazitäten. Dabei wird die Funktionalität gekoppelter Kapazitäten genutzt, Ladungsträger lateral zu transportieren. In Bild IV.8 ist veranschaulicht, wie eine gespeicherte Ladung durch externe Steuersignale von Kapazität zu Kapazität transportiert wird. Rechenoperationen mit den als Ladungen dargestellten Signalwerten sind jedoch nur begrenzt realisierbar. Detaillierte Beschreibungen zu den Verfahren der CCD-Technik sind in [Theu95] und [How79] zu finden.

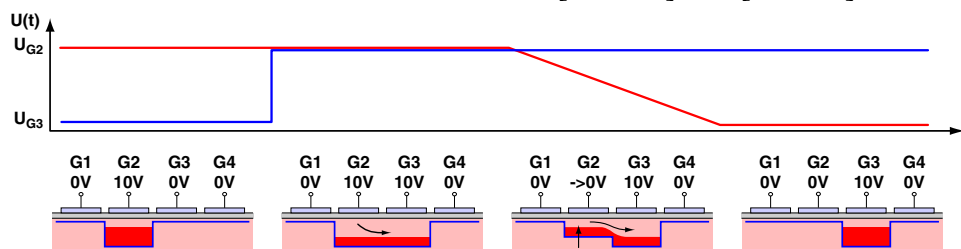


Bild IV.8 Prinzip des Ladungstransports bei CCD

Allgemein ist die CCD-Technologie weit entwickelt und weist aufgrund eines stark spezialisierten Fertigungsprozesses sehr geringe Verluste und Fehler auf. Es sind jedoch hohe Betriebsspannungen erforderlich, die mit hoher Verlustleistung verbunden sind. Außerdem können wegen der spezialisierten Fertigung keine anderen Bauelemente realisiert werden. Steuerschaltungen, Signalkonvertierung und -verarbeitung müssen daher mit zusätzlichen Schaltkreisen außerhalb des eigentlichen Sensors erfolgen (Bild IV.9). In jüngster Zeit wird an CCD/CMOS-Hybridssystemen gearbeitet [Hara04], mit denen die Vorteile von CMOS und CCD kombiniert werden sollen. In dieser Arbeit werden jedoch nur Photosensoren auf CMOS-Basis betrachtet.

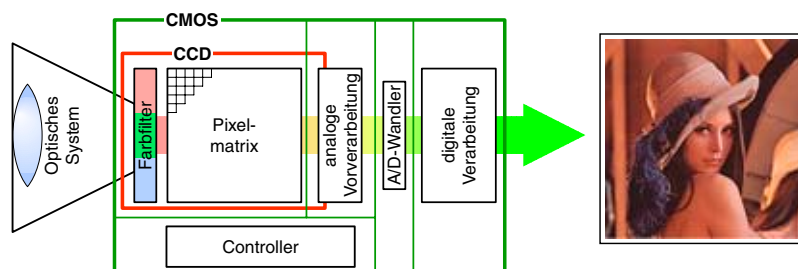


Bild IV.9 Blockschaltung und Funktionalität eines Kamerasystems mit CMOS und CCD

4.2 Passive Pixel (PPS)

Sensoren mit passiven Pixeln (*Passive Pixel Sensor* - PPS) stellen die einfachste Realisierung von CMOS-Photosensoren dar und bestehen nur aus einer Matrix von integrierenden Photodioden nach Abschnitt 3.3.5 mit zugehörigem Auswahlhalter, wie in Bild IV.10 skizziert. Zur praktischen Realisierung der Schalter sind MOS-Transistoren geeignet, die Auslese erfolgt in der Regel über einen Integrator. Zum Auslesen eines bestimmten Pixels wird der entsprechende

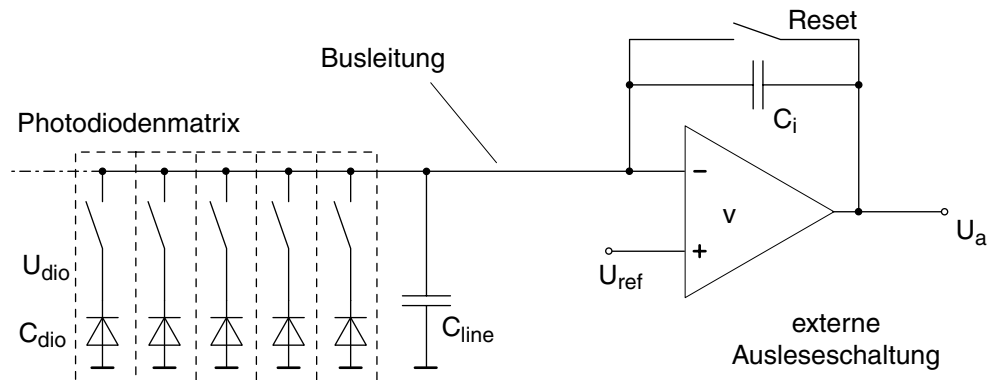


Bild IV.10 Prinzip eines passiven CMOS-Sensorsystems mit Integrationsauslese

Auswahlhalter aktiviert. Aufgrund der Differenz von U_{dio} und U_{ref} fließt daraufhin ein Strom, der die Sperrschichtkapazität der Diode auf U_{ref} rücksetzt. Die dazu nötige Ladung wird im Integrator gesammelt und kann am Ausgang als äquivalente Spannung abgenommen werden. Nach dem Auslesen wird der Auswahlhalter deaktiviert und der Integrator durch Schließen des Reset-Schalters zurückgesetzt. Die Ausgangsspannung ergibt sich bei idealer Betrachtung zu $U_a = U_{ref} - (U_{dio} - U_{ref}) \cdot C_{dio} / C_i$.

Der Vorteil der passiven Pixelfelder liegt im einfachen Aufbau. Da nur ein einziger Transistor notwendig ist, kann die Pixelfläche fast vollständig optisch aktiv ausgelegt werden. Allerdings ist die Geschwindigkeit der Auslese stark von der Kapazität und dem Widerstand der Busleitung beeinflusst. Die maximal realisierbare Leitungslänge und die Empfindlichkeit der Schaltung werden ebenfalls begrenzt. Die Integrationskapazität C_i entspricht infolge des Miller-Effektes einer eingangsseitigen Kapazität von $C_i' = (1 + v) \cdot C_i$. Durch die Leitungskapazität C_{line} ist die wirksame Integrationskapazität auf $C_i' / (C_i' + C_{line})$ begrenzt. Mit den modernen Technologien sind die parasitären Einflüsse der Leitungskapazität jedoch stark gesunken, wodurch auch passive Pixelstrukturen interessant werden.

Eine Wannen-Substratdiode von $10 \times 10 \mu\text{m}$ besitzt im allgemeinen eine wesentlich kleinere Kapazität als die Leitungskapazität, beispielsweise ca. 50 fF. Eine Busleitung von $0.2 \mu\text{m} \times 10 \text{ mm}$ und einem typischen Kapazitätsbelag von $75 \text{ aF}/\mu\text{m}^2$ bei modernen Technologien ergibt dagegen eine Leitungskapazität von 150 fF. Um einen sinnvollen Ausgangsspannungshub zu erreichen, muß die Integrationskapazität ähnlich groß wie die Diodenkapazität dimensioniert sein. Wird die Integrationskapazität C_i ebenfalls mit 50 fF ausgelegt, entspricht dies bei einer Verstärkung $v = 100$ ca. 5 pF an Eingang, deren wirksamer Anteil durch die Leitungskapazität auf 97% verringert wird. Wird die Integrationskapazität vergrößert, verringert sich zwar der Einfluß der Leitungskapazität, allerdings sinkt auch der Ausgangsspannungshub. Die Dimensionierung der Kapazitäten ist demnach primär vom gewünschten Arbeitsbereich abhängig, während die parasitären Effekte nur sekundär berücksichtigt werden können.

4.3 Aktive Pixel (APS)

Aktive Pixelschaltungen enthalten Verstärkerstrukturen innerhalb des Pixels, die durch die Entkopplung der Photoelemente von der externen Schaltung für verbesserte Empfindlichkeit und höhere Auslesegeschwindigkeiten sorgen [Fos97]. Aktive Pixelschaltungen wurden bereits 1968 vorgestellt [Nob68], konnten sich jedoch wegen des geringen Signal-Rausch-Abstandes nicht behaupten. Erst die technologischen Verbesserungen des Herstellungsprozesses in den letzten Jahren haben aktive Pixelschaltungen und damit CMOS-Bildsensoren ermöglicht. Der Begriff der aktiven Pixelsensoren (*Active Pixel Sensors* - APS) wurde dabei 1992 von Fossum eingeführt [Fos92]. Aktive Pixelschaltungen enthalten im wesentlichen drei Komponenten, das Photoelement, eine Reset- und eine Ausleseschaltung. Zusätzlich können auch andere Elemente zur Erweiterung der Funktionalität enthalten sein [NPNF97].

4.3.1 Resetschaltungen

Für die Funktion eines integrierenden Pixels ist es erforderlich, die Speicherkapazität zu Beginn der Integrationsphase auf einen Initialwert zu setzen. Dies geschieht durch eine Resetschaltung. Aus Gründen des geringen Flächenbedarfs besteht die übliche Realisierung aus einem NMOS-Transistor entsprechend Bild IV.11a. Die Nachteile dieser Variante bestehen im wesentlichen in der Einführung von kTC-Rauschen, auf das in Abschnitt 5.1.2 eingegangen wird, sowie in der Verringerung des am Ausgang möglichen Spannungspegels um eine Schwellspannung. Die Variation der Schwellspannungen der Resettransistoren in den einzelnen Pixelschaltungen hat weiterhin *Fixed-Pattern-Noise* zur Folge.

In aktuellen Technologien sind weiterhin stark angestiegene Leckströme der Transistoren zu verzeichnen, die sich negativ auf die Signalintegrität auswirken. Als platzsparendere Implementierung mit ähnlichen elektrischen Eigenschaften und geringeren Leckströmen wurden daher Resetdioden untersucht (Bild IV.11c). Dabei wird der Resetpegel U_{Reset} an eine in Flußrichtung geschaltene Diode gelegt. Die Speicherkapazität wird daraufhin auf $U_{\text{Reset}} - U_F$ aufgeladen, mit der Flußspannung der Diode U_F . Ist der Resetvorgang beendet, wird U_{Reset} auf ein Ruhepotential gelegt und die Diode gesperrt. Im Beispiel nach Bild IV.11 ist das Ruhepotential mit dem Massepotential identisch. Der dabei fließende geringe Sperrstrom addiert sich zum Dunkelstrom der Photodiode. Um den Taktdurchgriff über die Sperrschichtkapazität der Resetdiode gering zu halten, ist diese in der Realisierung möglichst klein auszuführen.

Ein PMOS-Transistor als Resetschalter ist im Vergleich zu einem NMOS-Transistor nicht von der Reduktion des möglichen Resetpegels um eine Schwellspannung und dessen Variation betroffen, da PMOS-Transistoren bei aktivem Reset im *strong-inversion*-Bereich arbeiten (Bild IV.11b). Ein PMOS-Transistor erfordert jedoch aufgrund der notwendigen getrennten n-Wanne wesentlich mehr Platz und verringert damit die nutzbare photoempfindliche Fläche. Ein neues Schaltungskonzept sieht daher vor, den Resettransistor zur Verringerung des Flächenbedarfs innerhalb der Photodiode zu integrieren, z.B. in der Wanne einer Wannen-Substrat-Photodiode entsprechend Bild IV.11d. Nachteilig ist dabei die Notwendigkeit von zusätzlichen Steuersignalen, die jedoch außerhalb der Pixelmatrix erzeugt werden können.

Die Einbettung des Resettransistors ist dabei sowohl für n-Wannen als auch für isolierte p-Wannen möglich, jedoch stehen letztere nur in einem *Tripel-Well*-Prozeß zur Verfügung. Bild IV.12 zeigt einen Realisierungsvorschlag mit den zum Betrieb notwendigen externen Schaltungen zur

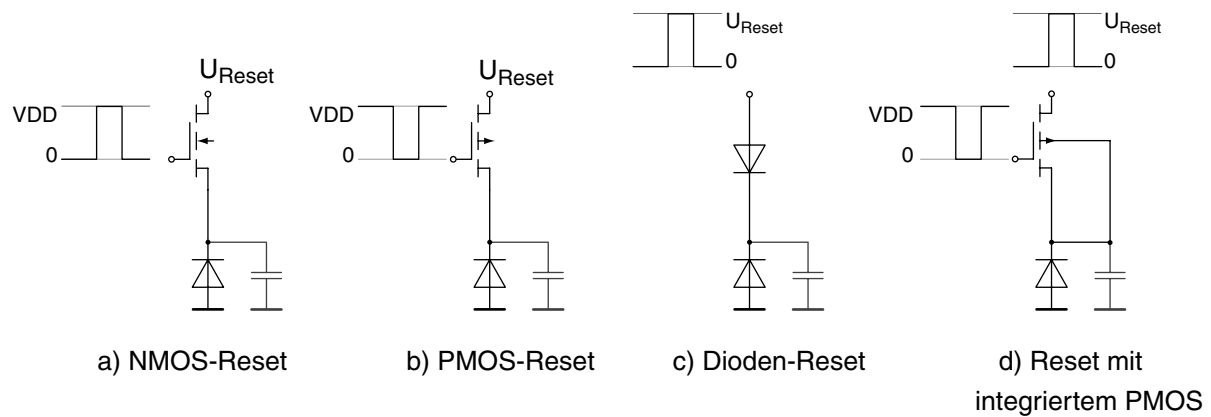


Bild IV.11 Schaltung und Ansteuerung für verschiedene Resetschaltungen
Signalgeneration.

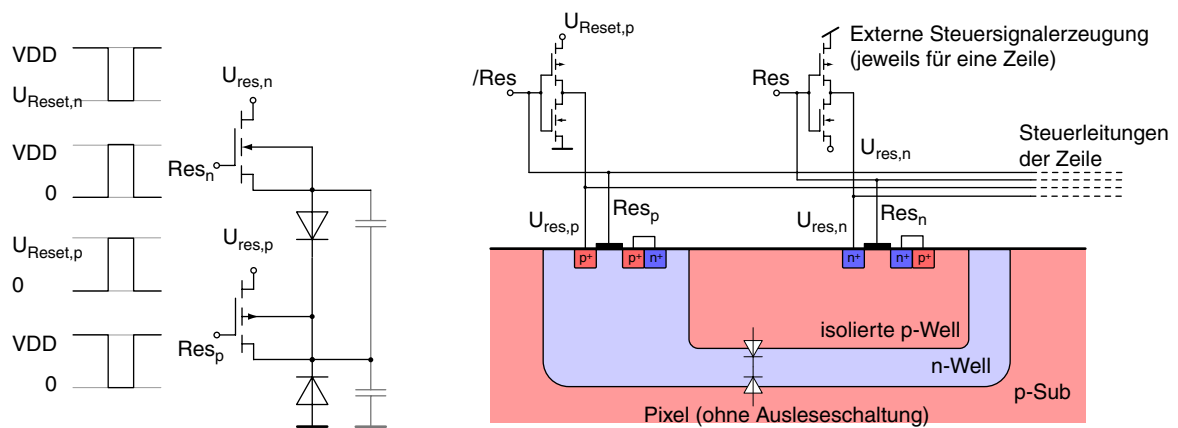


Bild IV.12 Eingebettete Resettransistoren als Schaltung und schematische Implantation

4.3.2 Einfache Ausleseschaltung

Die einfache Ausleseschaltung besteht aus einem Verstärker und einem Schalter. Der Verstärker sorgt für die Entkopplung der inneren Schaltung von den externen Kapazitäten und der Schalter dient zur Auswahl eines bestimmten Pixels an einer Busleitung. Die einfachste Realisierung des Verstärkers verwendet einen Sourcefolger [SG03]. Dabei ist jedoch anzumerken, daß der nutzbare Spannungshub am Ausgang um eine Schwellspannung reduziert ist.

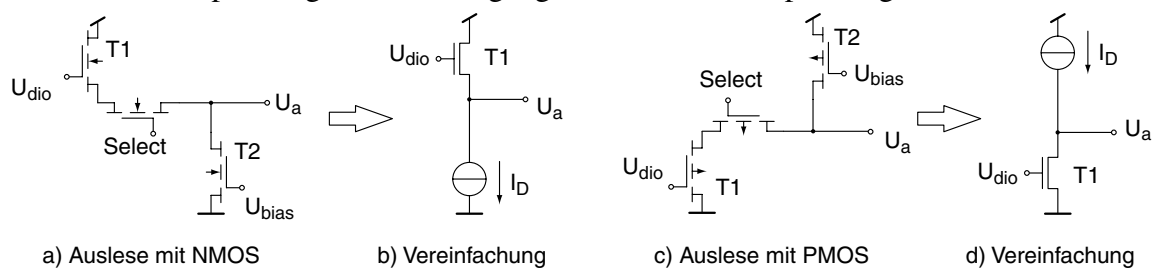


Bild IV.13 Ausleseschaltungen mit Sourcefolger

Betrachtet wird eine NMOS-Ausgangsstufe nach Bild IV.13a. Der Auswahltransistor T2 kann für die folgende Rechnung zur Vereinfachung entsprechend Bild IV.13b ignoriert werden, da er in Sättigung betrieben wird. Unter der Annahme, daß $U_a > U_{bias} - V_{T0}$ sei, ist T2 als Stromquelle idealisierbar. Für T1 gilt in diesem Fall:

$$I_D = \beta \cdot (U_{GS} - V_{T0})^2 = \beta \cdot (U_{dio} - U_a - V_{T0})^2 \quad (IV.18)$$

Umstellung ergibt:

$$U_a = U_{dio} - \left(V_{T0} + \sqrt{\frac{I_D}{\beta}} \right) \quad (\text{IV.19})$$

Die Ausgangsspannung des Sourcefolgers entspricht somit der Diodenspannung, reduziert um eine Schwellspannung und eine querstromabhängige Sättigungsspannung.

Das Arbeitsschema eines Bildsensors mit einfacher Auslese wird auch als rollender Shutter bezeichnet, da das Bild zeilenweise nacheinander gelesen und zurückgesetzt werden muß. Die Integration und das Auslesen der Bildzeilen erfolgt dabei für die einzelnen Zeilen zeitlich versetzt (Bild IV.14). Integrationszeit T_{int} und Auslesezeit T_Z pro Zeile bestimmen die Bildwiederholrate $1/T_{\text{Bild}}$ mit $T_{\text{Bild}} = T_{\text{int}} + T_Z$. Die maximale Bildwiederholrate ergibt sich bei Vernachlässigung der Reset-Zeit aus der minimalen Zeit, die notwendig ist, den gesamten Sensor auszulesen. Bei n Bildzeilen beträgt diese $n \cdot T_Z$ und die maximal verfügbare Integrationszeit ergibt sich folglich zu $(n - 1) \cdot T_Z$.

Nachteilig ist, daß der Beginn der Integrationszeiten für jede Bildzeile verschieden ist. Beispielsweise werden bewegte Objekte wegen des unterschiedlichen Integrationsbeginns auf zeilenweise verschiedene Bildorte abgebildet. Für höhere Ansprüche an die zeitliche Auflösung und die Entkopplung von Integration und Auslesevorgang ist daher ein Shutter unentbehrlich.

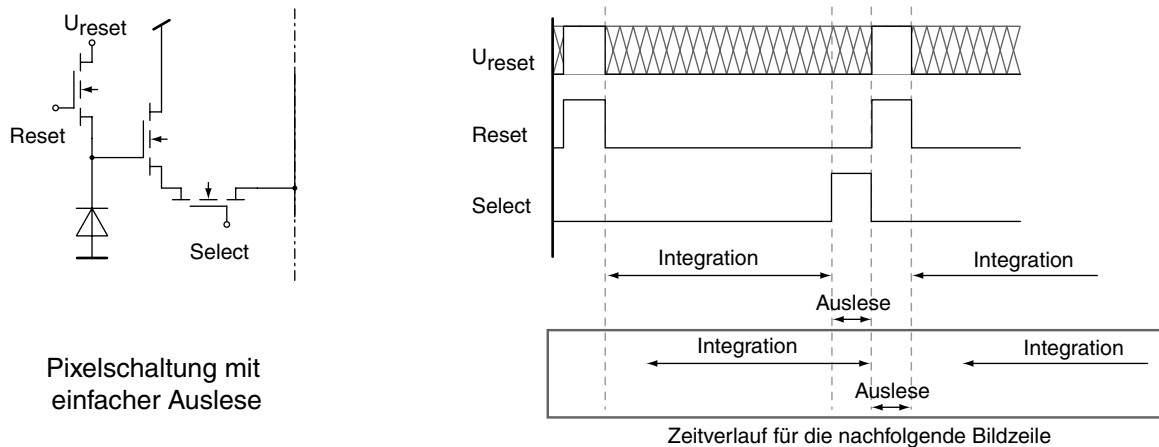


Bild IV.14 Pixelschaltung und Taktdiagramm für ein NMOS-Pixel mit einfacher Auslese

4.3.3 Ausleseschaltung mit elektronischer Shutter

Eine mögliche Ausleseschaltung mit elektronischem Shutter nach Bild IV.15 enthält eine zusätzliche Speicherkapazität. Diese kann als zusätzliches Bauelement realisiert sein oder aus der Gate-Kapazität der Auslestufe bestehen. Auf dieser Kapazität kann der aufgenommene Wert bis zum Auslesezeitpunkt zwischengespeichert werden und ermöglicht so eine gleichzeitige Integrationsphase für alle Pixel. Integration und Auslese verlaufen dabei unabhängig voneinander. Da Integrations- und Speicherkapazität nach dem Prinzip in Bild IV.15 nur gemeinsam zurückgesetzt werden können, müssen die beiden Vorgänge zeitlich nacheinander ablaufen. Die Bildwiederholrate ergibt sich damit zu $1/(T_{\text{int}} + n \cdot T_Z)$.

Wird für die Speicherkapazität ein eigenes Reset vorgesehen, ist ein weiterer Betriebsmodus realisierbar, bei dem Integration und Auslese gleichzeitig ablaufen. Die Bildwiederholrate wird in diesem Fall allein von der Integrationszeit bestimmt. Deren untere Schranke ergibt sich jedoch aus der Zeit $n \cdot T_Z$, die notwendig ist, um das gesamte Bild auszulesen. Eine andere Rea-

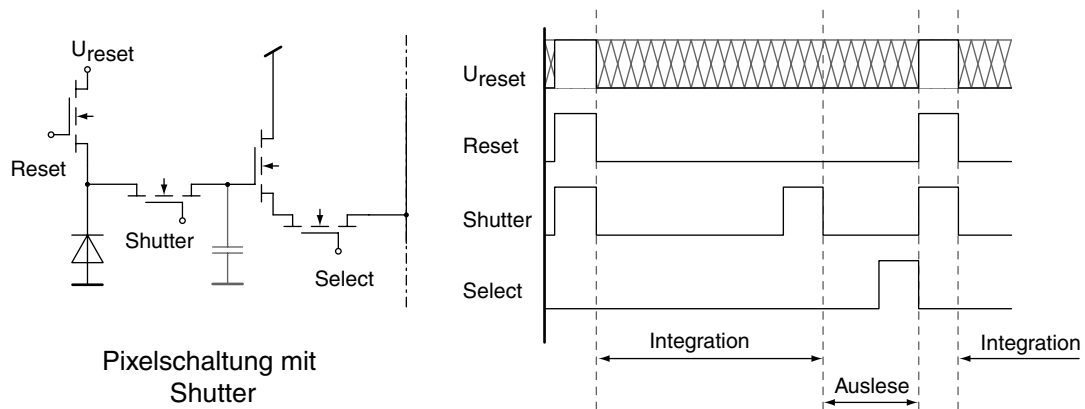


Bild IV.15 Pixelschaltung und Taktdiagramm für ein Pixel mit einfachem Shutter

lisierungsmöglichkeit mit den gleichen funktionalen Eigenschaften ergibt sich durch Einsatz eines Koppelverstärkers anstelle einer Resetschaltung der Speicherkapazität (Bild IV.16). Der Shutter dient in diesem Fall zum Abtasten der Ausgangsspannung des Koppelverstärkers.

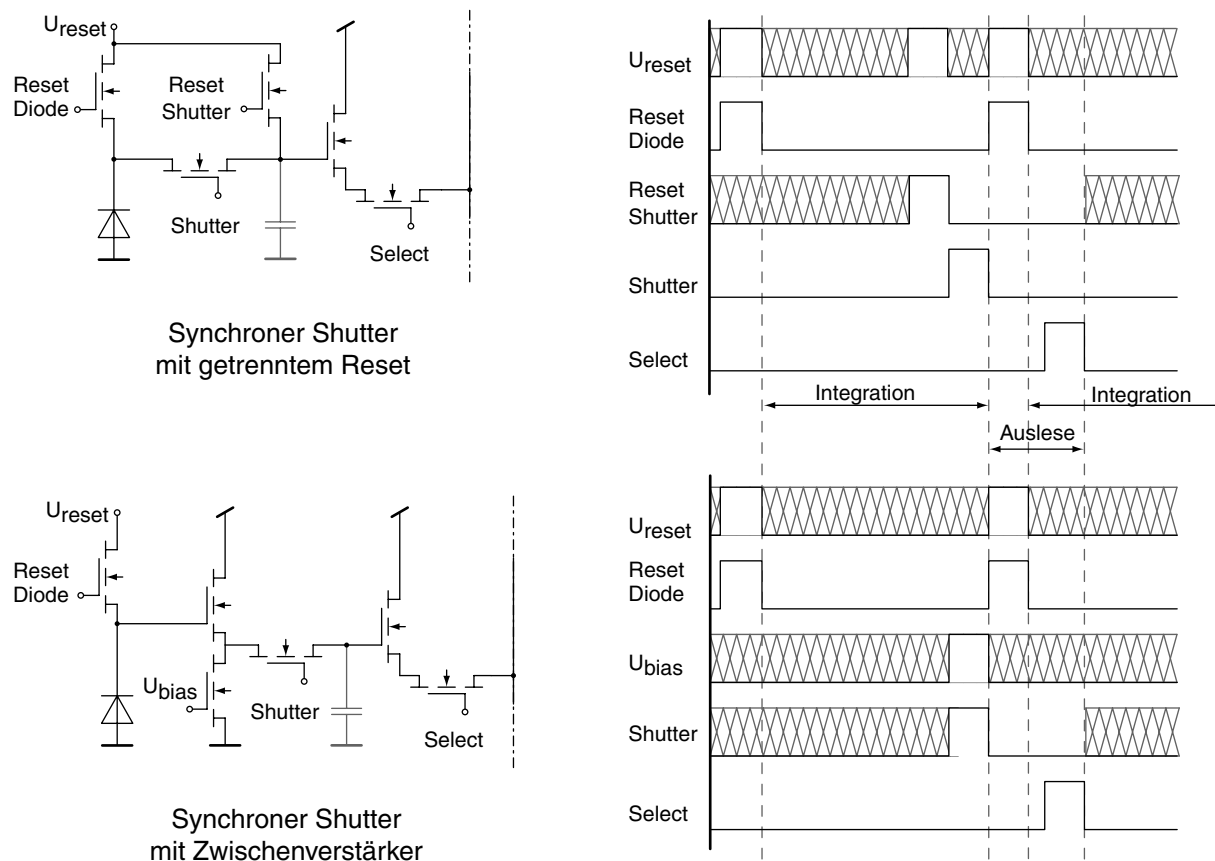


Bild IV.16 Pixelschaltung und Taktdiagramm für NMOS-Pixel mit synchronem Shutter

4.4 Logarithmische Pixel

Ein spezieller Typ eines aktiven Pixelsensors ist der logarithmische Photosensor [RD92]. Die Pixelzelle besteht aus einer Photodiode und drei Transistoren (Bild IV.17). Die Grundsaltung des logarithmischen Pixels arbeitet kontinuierlich und besitzt kein Reset.

Der Photostrom I_{ph} ist gleich dem Drain-Strom des Transistors T1, der im Subthreshold-Bereich arbeitet. Die Spannung über der Photodiode U_{ph} liegt somit nahe der Versorgungsspan-

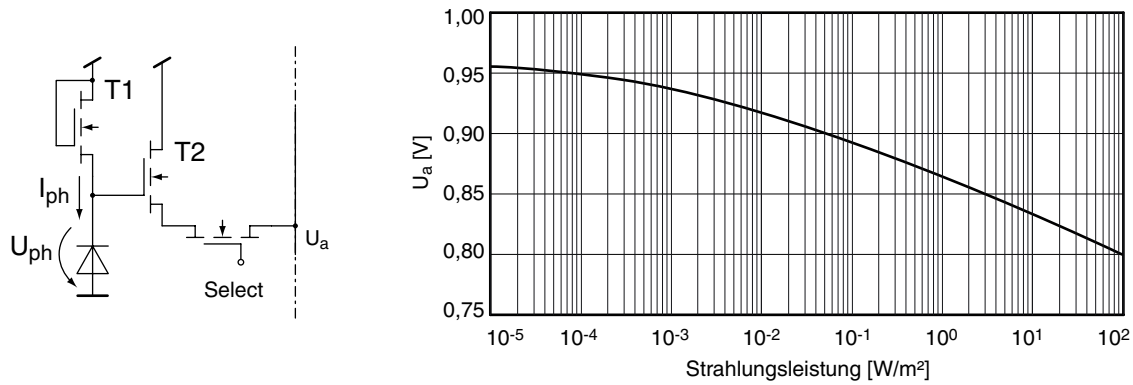


Bild IV.17 Pixelschaltung und Sensitivität für ein logarithmisches Pixel

nung V_{DD} und ergibt sich vereinfacht zu:

$$U_{ph} \approx V_{DD} - V_{T0} - U_T \cdot \ln\left(\frac{I_{ph}}{I_0}\right) \quad (\text{IV.20})$$

mit dem Sättigungsstrom I_0 und der Schwellspannung V_{T0} . U_T ist die Temperaturspannung, die bei 300 K typischerweise 26 mV beträgt. Die entscheidenden Nachteile dieser Pixelkonstruktion sind jedoch die starke Temperaturabhängigkeit, der äußerst geringe Spannungshub und das hohe *Fixed-Pattern-Noise*. Zur direkten Aufnahme von Farben ist die logarithmische Kennlinie ebenfalls nicht geeignet, da die notwendige Korrekturrechnung eine lineare Übertragungsfunktion erfordert (siehe auch Kapitel III, Abschnitt 3.2). Aufgrund der logarithmischen Übertragungsfunktion wird allerdings ein großer Dynamikbereich abgedeckt und der Einsatz des Sensors bei starken Beleuchtungsunterschieden innerhalb der Szene ermöglicht. Logarithmische Pixel sind daher im Bereich des maschinellen Sehens vielversprechend, da dort die Forderung nach einem hohen Dynamikbereich gegenüber exakter Farbaufnahme dominiert. Es existieren auch Schaltungsmodifikationen, die beispielsweise die logarithmische Pixelcharakteristik mit der von linearen Pixeln zu kombinieren [FHD00].

4.5 Digitale Pixel (DPS)

Digitale Pixelsensoren (DPS) stellen die neueste Entwicklung von Bildsensoren dar [GYF99] und enthalten Strukturen zur Analog-Digital-Wandlung in den einzelnen Pixeln (Bild IV.7). Die Architektur von digitalen Pixeln bietet gegenüber analogen Pixeln einige Vorteile, einschließlich der besseren Skalierbarkeit mit der CMOS-Technologie und geringerem Rauschen, da Teile der analogen Ausleseschaltung und die Spaltenverstärker entfallen [WGG02]. Durch die Integration von A/D-Wandler und Digitalspeicher in jedem Pixel werden zudem eine massiv parallele A/D-Wandlung und hohe Bildraten ermöglicht [YFG99], [KLLG01] (Bild IV.18). Nachteilig sind jedoch der hohe Aufwand für die Pixelschaltung, insbesondere für den Komparator und der dadurch stark reduzierte Anteil der photoaktiven Fläche an der Gesamtfläche des Pixels (Füllfaktor). Zudem haben Schwankungen der Bauelemente verstärkten Einfluß auf das Systemverhalten, da jedes Pixel eigene Wandlerstrukturen besitzt, die in ihren Parametern Abweichungen aufweisen. Digitale Pixel eignen sich jedoch besonders für Hochgeschwindigkeitsanwendungen, da im besonderen durch das hochparallele Wandlerprinzip sehr hohe Bildraten erreichbar sind.

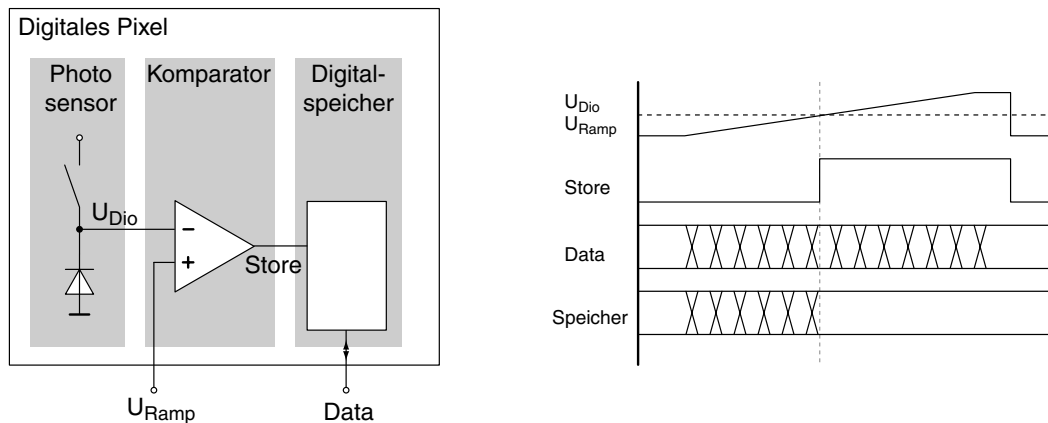


Bild IV.18 Prinzip eines digitalen Pixels nach [KLLG01]

5 Rauschen in Photosensoren

5.1 Rauschterme

Der Begriff Rauschen bezeichnet die statistischen Fluktuationen von Strömen und Spannungen, die in allen elektronischen Schaltungen auftreten. Die Ursachen dieses Rauschens liegen darin begründet, daß Ladung nicht kontinuierlich, sondern nur in Einheiten von Elementarladungen existiert. Das Rauschen stellt die fundamentale untere Grenze für die Meßgenauigkeit und die minimal verarbeitbare Signalstärke dar.

Bei Bildsensoren werden der nutzbare Dynamikumfang und die Bildqualität wesentlich durch das Rauschen bestimmt. Formal wird zwischen zeitlichem Rauschen (*random noise*) und örtlichem Rauschen (*pattern noise*) unterschieden. Zeitliches Rauschen ist unabhängig von der Schaltungstechnik und beschreibt die zufällige Variation der elektrischen Größen über Rauschleistungs dichtespektren [TFG01]. Die Angabe dieser Spektren kann dabei einseitig mit stets positiven Frequenzen f oder zweiseitig mit $f \in \mathbb{R}$ erfolgen. In der folgenden Betrachtung soll die zweiseitige Darstellung der Leistungs dichtespektren verwendet werden, die jeweils den halben Wert der Rauschleistung pro Frequenzintervall im Vergleich zur einseitigen Darstellung angibt. Örtliches Rauschen spiegelt sich in zeitlich unabhängigen Variationen der einzelnen Pixel eines Bildes wieder. Die Hauptkomponenten des örtlichen Rauschens sind das *Fixed-Pattern-Noise* (FPN) und Rauschen aufgrund von ungleichmäßigen Pixelempfindlichkeiten (*photo response non-uniformity*, PRNU). Während das *Fixed-Pattern-Noise* unabhängig von der Beleuchtung ist, wird die PRNU von Stärke und Wellenlänge der Beleuchtung beeinflusst. Zur Beschreibung des örtlichen Rauschens wird allgemein eine Normalverteilung mit Mittelwert und Varianz angenommen, wobei der Mittelwert oft Null ist. Die Hauptursache von FPN bei CMOS-Bildsensoren mit aktiven Pixelschaltungen liegt u.a. in der Variation der Transistorschwellspannung begründet. Dies betrifft die Resetstufe, den Pixelverstärker und auch die Schaltungen außerhalb der Pixelmatrix, die zur Weiterverarbeitung der Meßwerte dienen. Die Variation der Schwellspannung ist allgemein von der Dimensionierung der Transistoren abhängig und verringert sich allgemein mit größeren Strukturen. Geringe Transistorabmessungen sind jedoch für eine hohe Integrationsdichte wünschenswert. Ungleichmäßige Pixelempfindlichkeiten werden von lokalen Variationen des Absorptionskoeffizienten, lokalen Störstellen und Nichtlinearitäten der Ausleseschaltung verursacht. Zu den Auswirkungen des Rauschens in Bildern auf die Wahrneh-

mung existieren aktuelle Forschungen [WS04], die zwar auf die Anwendung in verlustbehafteter Kompression ausgelegt, jedoch auch für das Sensordesign interessant sind.

5.1.1 Thermisches Rauschen

Thermisches Rauschen tritt an allen ohmschen Bauteilen in der Schaltung auf und wird allgemein als weißes Rauschen mit konstantem Leistungsdichtespektrum modelliert. Da dies jedoch einer unendlichen Rauschleistung entspricht, kann weißes Rauschen in physikalischen Systemen nicht vorkommen. Das Modell ist allerdings für Rauschprozesse geeignet, deren Leistungsdichtespektren im interessierenden Frequenzbereich nahezu konstant sind. Dies wird allgemein mit einer konstanten Rauschleistungsdichte und einem idealen Tiefpaß der Bandbreite B beschrieben. Die rausch-äquivalente Bandbreite B wird dabei so bestimmt, daß die Rauschleistungen des modellierten Systems mit der Übertragungsfunktion $A(\omega)$ und der Idealisierung mit idealem Tiefpaß übereinstimmen:

$$A(0)^2 \cdot 4\pi \cdot B = \int_{-\infty}^{\infty} |A(\omega)|^2 \cdot d\omega \quad (\text{IV.21})$$

$$B = \frac{1}{4\pi \cdot A(0)^2} \cdot \int_{-\infty}^{\infty} |A(\omega)|^2 \cdot d\omega$$

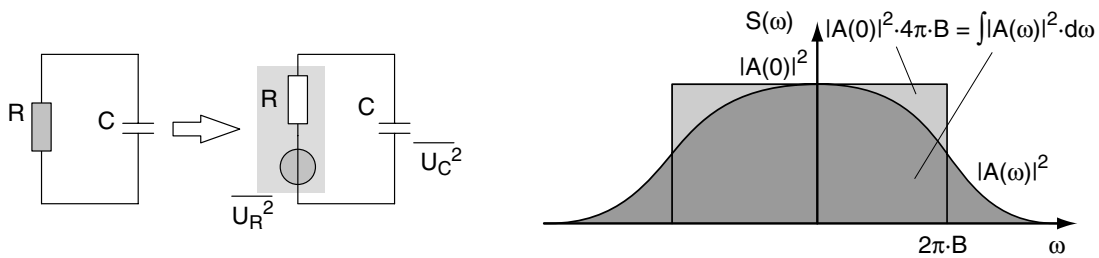


Bild IV.19 Modellierung des thermischen Rauschens

Für einen RC-Tiefpaß (Bild IV.19) mit der Übertragungsfunktion $A(\omega) = 1 / (1 + j\omega RC)$ ergibt sich die rausch-äquivalente Bandbreite nach Gleichung (IV.21) somit zu $B = 1 / (4RC)$. Nach dem Gleichverteilungssatz der statischen Physik nimmt ein System im thermischen Gleichgewicht pro Freiheitsgrad die mittlere thermische Energie $kT/2$ mit der absoluten Temperatur T und der Boltzmann-Konstante k an [Hau99]. Daraus ergibt sich:

$$\frac{1}{2} \cdot C \cdot \overline{U_C^2} = \frac{kT}{2} \quad \text{b.z.w.} \quad \overline{U_C^2} = \frac{kT}{C} \quad (\text{IV.22})$$

Die äquivalente Rauschspannung $\overline{U_C} = \sqrt{kT/C}$ am Ausgang des RC-Tiefpasses ist somit unabhängig von der Dimensionierung des Widerstandes R . Das Leistungsdichtespektrum der äquivalenten Rauschspannungsquelle des thermisch rauschenden Widerstandes für das am Eingang angenommene weiße Rauschen ergibt sich folglich zu $S_{U,R}(\omega) = 2kTR$, beziehungsweise $S_{I,R}(\omega) = 2kT/R$ für die äquivalente Rauschstromquelle. Ein rauschender Widerstand kann somit nach Gleichung (IV.22) durch einen idealen Widerstand und eine Rauschspannungs- respektive Rauschstromquelle mit Effektivwerten von $\overline{U_R} = \sqrt{4kTRB}$ respektive $\overline{I_R} = \sqrt{4kTB/R}$ beschrieben werden.

Ebenfalls existiert mit der Ladung $Q = C \cdot \overline{U_C} = \sqrt{kTC}$ eine untere Grenze für die Detektierbarkeit von Ladungsträgern in SC-Schaltungen¹⁷. Für eine Kapazität von 1 pF und Raumtemperatur beträgt diese Grenze $Q = 6.44 \cdot 10^{-17} \text{ A}\cdot\text{s}$.

17. SC = Switched Capacitor, geschaltete Kapazitäten

5.1.2 kTC-Rauschen

Auch eine Pixelzelle wird vom thermischen Rauschen eines Tiefpasses beeinflusst. Der Reset-Vorgang (siehe auch Kapitel IV, Abschnitt 4.3.1) kann vereinfacht als Laden einer Kapazität C über den ohmschen Widerstand eines MOSFET oder einer Diode betrachtet werden. Die äquivalenten Rauschspannung $\overline{U_{kTC}} = \sqrt{kT/C}$ wird in der Photosensorik zu Vergleichszwecken üblicherweise in der Anzahl der akkumulierten Elektronen dargestellt und ergibt sich zu:

$$n_{e, kTC} = \frac{C}{e} \cdot \frac{\sqrt{kT}}{\sqrt{C}} = \frac{\sqrt{kTC}}{e} \quad (IV.23)$$

Aufgrund der Termini wird das Rauschen des Resetvorgangs häufig als kTC-Rauschen bezeichnet. Bei Raumtemperatur kann vereinfacht mit $n_{e, kTC} \approx 400 \sqrt{C/pF}$ gerechnet werden. Für die Sperrschichtkapazität einer Photodiode von beispielsweise 50 fF ergeben sich etwa 90 Elektronen, die vom kTC-Rauschen verursacht werden. Zur Verringerung des kTC-Rauschen existieren viele Vorschläge, effektiv sind jedoch nur die Verringerung von Temperatur und Kapazität. Ersteres ist für Spezialanwendungen mit gekühlten Sensorschaltkreisen realisierbar, für den Massenmarkt jedoch nicht praktikabel. Die Kapazität ist wiederum im wesentlichen von technologischen Randbedingungen abhängig und kann im Entwurfsprozeß nur durch Wahl der aktiven Fläche beeinflusst werden. Eine Verkleinerung der Photodiode zur Kapazitätsreduktion ist jedoch konträr zu der Forderung, möglichst viele Photonen einzufangen. Das kTC-Rauschen ist somit ein Rauschterm, der nur mittelbar im Entwurfsfluß beeinflusst werden kann.

5.1.3 Schrotrauschen

Schrotrauschen wird durch das stochastische Eintreffen der Photonen bzw. die stochastische Ladungsträrgbewegung verursacht. Charakteristisch für das Schrotrauschen ist, daß sich das zeitliche Rauschsignal aus zufällig verteilten, statistisch unabhängigen Einzelimpulsen zusammensetzt, die mit einer Rate ν auftreten. Die Anzahl der innerhalb eines Zeitintervalls auftretenden Impulse ist unter diesen Bedingungen Poisson-verteilt. Mittelwert und Varianz der Impulsanzahl N innerhalb einer Zeit T sind demnach:

$$E(N) = \nu \cdot T \quad \sigma_N^2 = \nu \cdot T \quad (IV.24)$$

Ist der zeitliche Verlauf eines Impulses durch $g(t)$ mit der Transformierten $G(\omega)$ beschrieben, so ergibt sich die Leistungsdichte des Schrotrauschens nach dem Theorem von Carlson zu:

$$S(\omega) = \nu \cdot |G(\omega)|^2 \quad (IV.25)$$

Ein elektrischer Strom I ist beschrieben als Fluß einer Anzahl von Ladungsträgern der Elementarladung e pro Zeiteinheit. Das Leistungsdichtespektrum des Schrotrauschens des Stromflusses ist demnach $S_{Schrot}(\omega) = e \cdot I \cdot |G(\omega)|^2$. Wird weiterhin die Impulsform $g(t) = \delta(0)$ als Dirac-Impuls idealisiert, so ergibt sich die als Schottky-Formel bekannte Beziehung:

$$S_{Schrot}(\omega) = e \cdot I \quad (IV.26)$$

Das Schrotrauschen weist somit eine dem Stromfluß proportionale Rauschleistungsdichte auf und kann wie das thermische Rauschen als weißes Rauschen mit der Bandbreite B modelliert werden. Der äquivalente Rauschstrom ist somit $\bar{I}_{Schrot} = \sqrt{2e \cdot I \cdot B}$, die Rauschladung bei einer Integrationszeit t_{int} berechnet sich mit $\bar{Q}_{Schrot} = \sqrt{2e \cdot I \cdot t_{int}}$ und die Anzahl der Rauschelektronen mit $n_{e, Schrot} = \sqrt{2I \cdot t_{int}} / e$. Der Strom I setzt sich im Falle einer Photodiode

aus dem Dunkelstrom I_0 und dem Photostrom I_{ph} zusammen. Praktische Werte für das Schrotrauschen in Photodioden liegen bei einigen hundert äquivalenten Elektronen.

Das Schrotrauschen kann im Entwurfsfluß nicht direkt beeinflußt werden. Der Dunkelstrom wird im wesentlichen von der Technologie bestimmt. Durch Vermeidung von spezifischen Layoutstrukturen (z.B. Ecken) können jedoch Störstellen und damit auch der Dunkelstrom in Grenzen verringert werden. Die Ströme lassen sich zwar prinzipiell durch Verkleinerung der Diodenfläche verringern, ein hoher Photostrom ist jedoch für eine adäquate Messung wichtiger, so daß eine Photodiode so groß wie möglich dimensioniert wird.

5.1.4 Photonenrauschen

Photonenrauschen ist eine Spezialform des Schrotrauschens, mit dem das Rauschen der stochastisch eintreffenden Photonen beschrieben wird. Es stellt eine fundamentale Einschränkung bei der Messung von Strahlung dar, da es der zu messenden Größe selbst innewohnt. Detektoren, deren Eigenrauschen kleiner als das Photonenrauschen ist, werden als *photon-noise-limited* bezeichnet.

Für einen linearen integrierenden Photosensor gilt, daß die Anzahl der akkumulierten Elektronen proportional der Anzahl der eintreffenden Photonen ist. Bei einer Bestrahlungsstärke E ist der Erwartungswert¹⁸ der Anzahl der Photonen mit Wellenlänge λ , die innerhalb einer Zeit T_{int} auf einer Fläche A eintreffen, bestimmt durch $E(N) = N = T_{int} \cdot A \cdot E \cdot \lambda / (h \cdot c)$. Da diese Anzahl Poisson-verteilt ist, ist auch deren Varianz $\sigma_N^2 = E(N) = N$. Für den erreichbaren Signal-Rauschabstand (SNR¹⁹) ergibt sich damit:

$$SNR_{dB} = 20dB \cdot \log\left(\frac{E(N)}{\sigma_N}\right) = 10dB \cdot \log(N) \quad (IV.27)$$

Der maximal mögliche Signal-Rauschabstand ist somit von der Bestrahlungsstärke E , der Integrationszeit T_{int} und der wirksamen Sensorfläche A begrenzt.

Einer Beleuchtungsstärke von 500 lx entspricht für Licht von 555 nm Wellenlänge einer Bestrahlungsstärke von $0.735 \text{ W} \cdot \text{m}^{-2}$ und somit $2.05 \cdot 10^{18} \text{ Photonen} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. Für einen Sensor mit $10 \mu\text{m} \times 10 \mu\text{m}$ aktiver Fläche und 10 ms Belichtungszeit ergeben sich damit $2.05 \cdot 10^6$ eintreffende Photonen und ein maximal mögliches SNR von 63 dB. In der Betrachtung nach Abschnitt 5.1.3 entspricht das Photonenrauschen dem Schrotrauschen, das vom Photostrom I_{ph} verursacht wird. Zur Ermittlung der Rauschladung ist die Kenntnis der Quanteneffizienz erforderlich, mit der bestimmt ist, wie viele Photonen zur Ladungsträgergeneration beitragen. Wird diese mit 0.5 angenommen, so generieren die eintreffenden Photonen bei Vernachlässigung von Randeffekten einen Photostrom von 16 pA. Die äquivalente Rauschladung des Photonenrauschens beträgt in diesem Fall 226 aA·s oder ca. 1400 Rauschelektronen.

5.1.5 1/f-Rauschen und Flicker-Rauschen

In allen elektronischen Elementen findet sich eine weitere vom fließenden Strom abhängige Rauschkomponente, deren spektrale Leistungsdichte proportional zu $|f|^{-\alpha}$ ist. Der Parameter α nimmt dabei in aller Regel Werte an, die nur wenig größer als Eins sind. Das sogenannte 1/f-

18. Der Erwartungswert einer Größe X sei mit $E(X)$ beschrieben und nicht mit der Bestrahlungsstärke E zu verwechseln.

19. SNR - *Signal to Noise Ratio*, Signal-Rausch-Abstand

Rauschen ist damit von einer spektralen Leistungsdichte gekennzeichnet, die um 3 dB pro Oktave abnimmt und ist auch in anderen natürlichen Phänomenen zu beobachten. 1/f-Rauschen unterscheidet sich formal von dem Rauschphänomen, das in der Literatur als Flicker- oder Funkel-Rauschen²⁰ bezeichnet wird, dessen spektrale Leistungsdichte mit etwa 6 dB pro Oktave abfällt. 1/f-Rauschen tritt bei allen aktiven und einigen passiven Bauteilen, wie beispielsweise Karbon-Widerständen, auf. Es wird durch Haftstellen verursacht, die zufällig Ladungsträger festhalten und wieder freigeben. In Halbleitern entstehen diese Haftstellen durch Oberflächeneffekte, Verunreinigungen und Defekte im Kristall. Als Beschreibung des 1/f-Rauschens wird eine eingangsbezogene spektrale Leistungsdichte der Form $S_{U, 1/f}(\omega) = K \cdot f^{-\alpha}$ verwendet. Die Integration der Leistungsdichte über ein Frequenzintervall $[f_u, f_o]$ für $\alpha = 1$ ergibt:

$$\bar{u}^2 = K \cdot \ln(f_o/f_u) = K \cdot \ln 10 \cdot \log(f_o/f_u) \quad (\text{IV.28})$$

Sowohl für $f_o \rightarrow \infty$ als auch für $f_u \rightarrow 0$ geht die Rauschleistung gegen unendlich. Während für die obere Grenze aufgrund der allgemeinen Tiefpaßcharakteristik eines realen Systems eine Näherung gefunden werden kann, ist die Angabe einer unteren Grenzfrequenz formal nicht möglich. Allgemein kann zumindest die reziproke Betriebsdauer des praktischen Gerätes als untere Grenze angenommen werden. Bei niedrigen Frequenzen stellt das 1/f-Rauschen den dominanten Rauschterm dar. Für den Entwurf rauscharmer MOS-Schaltungen ist die Berücksichtigung des 1/f-Rauschens wichtig, da es besonders bei kleinen Transistoren bis in den MHz-Bereich gegenüber dem thermischen Rauschen überwiegen kann [TG01].

5.1.6 Quantisierungsrauschen

Das Quantisierungsrauschen ist eine weitere Störung, die durch die Quantisierung eines Signals hervorgerufen wird. Bei Annahme einer Gleichverteilung des zu quantisierenden Signals und Vollaussteuerung ergibt sich der Effektivwert des äquivalenten Rauschsignals zu $\bar{S}_Q = S/2^n$. In der Praxis wird die maximal sinnvolle Bitbreite bei der Digitalisierung danach bestimmt, daß das Quantisierungsrauschen kleiner als das Rauschen des Analogsignals bei gleichen Größenordnungen ist und damit vernachlässigt werden kann.

5.2 Rauschmodelle

5.2.1 Rauschmodell einer Photodiode

Bild IV.20 zeigt das Rauschersatzschaltbild einer Photodiode. Die Kapazität C_D entspricht der Sperrschichtkapazität nach Gleichung (IV.15). Der rauschfreie differentielle Ausgangsleitwert

20. Das Funkelrauschen wurde erstmalig 1925 von Johnson [John25] experimentell und 1926 von Schottky [Schot26] durch theoretische Analyse von Meßresultaten beschrieben. Der Effekt wurde bei der Emission von Elektronen aus Glühkathoden beobachtet und von Schottky durch Funkeln (eng.: flicker) an der Oberfläche der Kathode erklärt. Durch Übertragung des englischen Begriffes „flicker noise“ entstand die deutsche Bezeichnung Flicker-Rauschen, die äquivalent zu Funkelrauschen gebräuchlich ist. Schottky bestimmte aus Johnsons Messungen zwei unterschiedliche spektrale Rauschleistungsdichten für unterschiedliche Kathodenmaterialien, deren Verläufe proportional zu f^{-2} und zu $(f_o^2 + f^2)^{-1}$ sind. In der Elektroakustik werden jedoch sowohl 1/f-Rauschen als auch Funkelrauschen als rosa Rauschen bezeichnet.

g_D bestimmt sich durch Ableitung der allgemeinen Diodengleichung zu:

$$g_D = \frac{\delta}{\delta U} \cdot I_D = \frac{I_S}{U_T} \cdot e^{U/U_T} \quad (\text{IV.29})$$

Die spektrale Leistungsdichte der Rauschstromquelle i_D ergibt sich aus dem Schrotrauschen der Diode: $S_I(\omega) = e \cdot [I_S \cdot (1 + e^{U/U_T}) + I_{ph} + I_{dark}]$. Da für Sperrspannungen $-U \gg U_T$ der Exponentialterm vernachlässigt werden kann, folgt die vereinfachte Beziehung mit $S_I(\omega) = e \cdot [I_S + I_{ph} + I_{dark}]$.

5.2.2 Rauschmodell eines MOS-Transistors

Die Rauschanteile eines MOS-Transistors in starker Inversion setzen sich aus dem thermischen Rauschen und $1/f$ -Rauschen zusammen. Bei niedrigen Frequenzen überwiegt in der Regel der Anteil des $1/f$ -Rauschens. Die verschiedenen Beiträge zum Rauschen können eingangs- oder ausgangsseitig zusammengefaßt werden, wobei unkorrelierte Rauschbeiträge addiert werden können. Die Beziehung der Leistungsdichtespektren von eingangsseitiger Rauschspannungsquelle und ausgangsseitigen Rauschstromquelle lautet $S_{I,a}(\omega) = g_m^2 \cdot S_{U,e}(\omega)$.

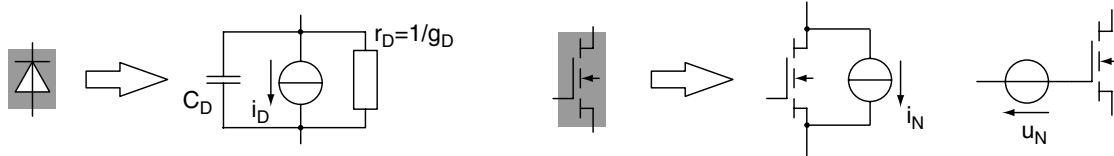


Bild IV.20 Rauschersatzschaltungen für Diode und MOS-Transistor

Für einen Transistor in Sättigung ergibt sich die Rauschleistungsdichte des thermischen Rauschens näherungsweise zu:

$$S_{u,th}(\omega) = 4kT \cdot \gamma \cdot \frac{1}{g_m} \quad (\text{IV.30})$$

In den Parameter γ geht u.a. die Transistorgeometrie ein, für Langkanal-Transistoren gilt allgemein $\gamma \approx 0.6$ und für Kurzkanal-Transistoren $\gamma \approx 2.5$ [TBK96]. Im ohmschen Bereich gilt:

$$S_{u,th}(\omega) = 4kT \cdot \gamma \cdot \frac{1}{g} \quad (\text{IV.31})$$

mit dem Leitwert $g = g_{DS} = \mu \cdot C'_{ox} \cdot W/L \cdot (U_G - V_{T0} - U_D)$ und der effektiven Oxidkapazität C'_{ox} und den Transistorabmessungen W und L . Das $1/f$ -Rauschen ist im Gegensatz zum thermischen Rauschen unabhängig vom Arbeitspunkt und kann mit

$$S_{u,1/f}(\omega) = \frac{k_f}{C'_{ox} \cdot W \cdot L} \cdot \frac{1}{f} \quad (\text{IV.32})$$

beschrieben werden, wobei der Parameter k_f stark technologieabhängig ist.

5.3 Rauschen von Pixelschaltungen

Das Rauschen am Ausgang einer Pixelschaltung läßt sich durch Superposition aller unabhängigen Rauschprozesse ermitteln und soll hier am Beispiel einer einfachen Pixelschaltung nach Bild IV.14 demonstriert werden. Als wichtigste Terme sollen das kTC-Rauschen der Reset-schaltung sowie das thermische und $1/f$ -Rauschen der Ausgangsstufe in Relation zum Photonenrauschen betrachtet werden. Zur Vergleichbarkeit werden die Rauschterme als äquivalente Rauschelektronen für ein Pixel mit $10 \mu\text{m} \times 10 \mu\text{m}$ aktiver Fläche, 10 ms Belichtungszeit sowie einer Quanteneffizienz η von 0.5 und Vernachlässigung aller Randeffekte angegeben.

Für das kTC-Rauschen ergeben sich bei 50 fF Diodenkapazität nach Abschnitt 5.1.2 $n_{\text{kTC}} = 90$ Rauschelektronen. Das thermische Rauschen und das 1/f-Rauschen der Ausgangsstufe kann ebenfalls auf eine äquivalente Rauschladung an deren Eingang rückgerechnet werden. Für das thermische Rauschen der im ohmschen Bereich arbeitenden Auslesetransistoren ergibt sich nach Gleichung (IV.31) bei Annahme von Kurzkanal-Transistoren und einem Leitwert von $85 \mu\text{S}$ eine eingangsseitige Rauschspannung von nur 312 nV, die vernachlässigt werden kann. Die Angabe des 1/f-Rauschen ist ohne Kenntnis des technologieabhängigen Parameters k_f nicht möglich. In der Literatur existieren allerdings Angaben zu andere Technologien, aus denen sich Näherungen für sub-100nm-Technologien ermitteln lassen. In [Tian00] und [TG01] sind für Transistoren mit $1 \mu\text{m} \times 1 \mu\text{m}$ Kanalfläche äquivalente Rauschspannungen von ca. 200 μV angegeben. Eine Extrapolation ergibt für die in dieser Arbeit betrachteten Minimaltransistoren in sub-100nm eine äquivalente Rauschspannung von ca. 680 μV , die einer äquivalenten Rauschladung von 34 aA·s bzw. $n_{1/f} = 210$ Rauschelektronen entspricht. Das vom Sensor verursachte Rauschen beläuft sich folglich auf $\sqrt{n_{1/f}^2 + n_{\text{kTC}}^2} = 229$ Elektronen.

Eine Messung von Strahlung ist nur oberhalb dieser Grenze möglich, die bei einer angenommenen Quanteneffizienz η von 0.5 und Licht mit 555 nm Wellenlänge einer Beleuchtungsstärke von 0.112 lx entspricht. Das Sensorrauschen überwiegt dabei gegenüber dem Photonenrauschen nach Abschnitt 5.1.4 bis zu einer Beleuchtungsstärke von 26 lx. Die Obergrenze der meßbaren Strahlung wird vom maximal auswertbaren Spannungshub auf der Diodenkapazität bestimmt. Für 2 V Hub auf 50 fF ergeben sich dahingehend $0.62 \cdot 10^6$ Photoelektronen, die einer Beleuchtungsstärke von ca. 300 lx bei 555 nm Wellenlänge entsprechen. Dabei ist nur im Bereich zwischen 26 und 300 lx eine vom Sensorrauschen unabhängige Strahlungsmessung möglich.

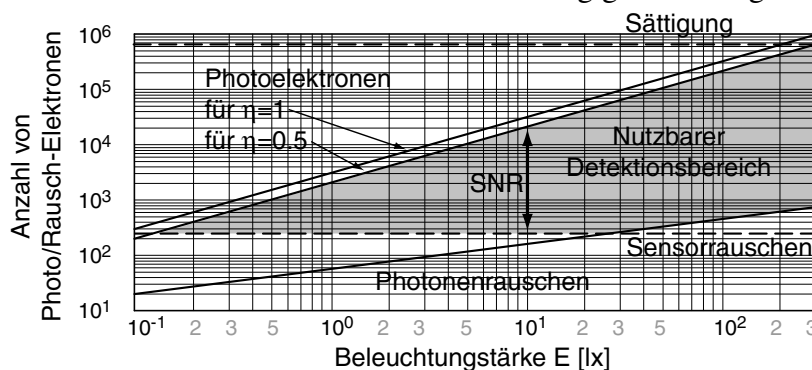


Bild IV.21 Übertragungscharakteristik eines Sensors

5.4 Methoden zur Rauschunterdrückung

Die Elemente eines Bildsensors liefern unterschiedliche Beiträge zum Gesamtrauschen und können weitgehend unabhängig voneinander betrachtet werden. Viele Rauschparameter sind dabei jedoch technologisch bestimmt und können im Entwurfsprozeß nicht direkt beeinflusst werden. Die verfügbaren Optimierungsparameter beschränken sich auf die Auswahl geeigneter Bauteile und deren Dimensionierung. Dies betrifft insbesondere die Abmessungen der Transistoren. Für den Entwurf rauscharmer Schaltungen ist daher eine Analyse des Einflusses der einzelnen Rauschanteile der Bauelemente erforderlich. Dies kann beispielsweise durch eine Varianzanalyse oder Monte-Carlo-Analyse erfolgen [Grau03]. Für komplexere Strukturen haben sich auch Systemsimulationen als geeignet erwiesen [HGG01]. Aus den Ergebnissen können die Elemente mit den größten Störeinflüssen gezielt optimiert werden, während unkritische

Bauteile nach anderen Kriterien, wie beispielsweise minimalem Flächenbedarf, entworfen werden können. Zur Verringerung des $1/f$ -Rauschen eignet sich, soweit möglich, eine Vergrößerung der Kanalfäche der Transistoren, während das kTC -Rauschen und Schrotrauschen nicht durch die Transistordimensionierung beeinflusst werden können.

Das zeitliche Rauschen in Bildern kann zudem unabhängig von der Realisierung durch Mittelung über mehrere Aufnahmen der gleichen Szene verringert werden. Dies ist jedoch nur bedingt bei der Aufnahme von bewegten Objekten anwendbar. Eine weitere Methode zur Rauschunterdrückung besteht im Einsatz der korrelierten Doppelabtastung.

5.4.1 Korrelierte Doppelabtastung

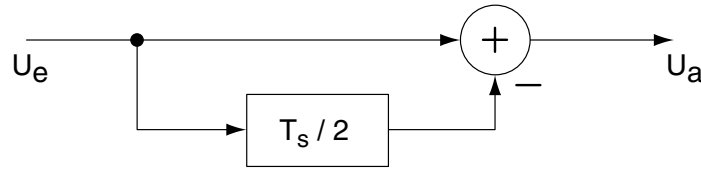


Bild IV.22 Blockschaltbild der Korrelierten Doppelabtastung

Die korrelierte Doppelabtastung (CDS²¹) beruht auf der Differenzbildung zwischen Signal- und Vergleichswert. Dadurch werden signalunabhängigen Fehler entfernt, die in beiden Termen enthalten sind, wie beispielsweise Offsets. Weiterhin werden niederfrequente Rauschanteile unterdrückt. Die Übertragungsfunktion des CDS-Systems nach Bild IV.22 im Zeitbereich und deren z -Transformierte lauten:

$$u_a(n \cdot T_s) = u_e(n \cdot T_s) - u_e\left(\left(n - \frac{1}{2}\right) \cdot T_s\right) \quad (\text{IV.33})$$

$$U_a(z) = G(z) \cdot U_e(z) = U_e(z) - z^{-1/2} \cdot U_e(z) = (1 - z^{-1/2}) \cdot U_e(z) \quad (\text{IV.34})$$

Eine Substitution von $z = e^{j\omega \cdot T_s}$ in der Übertragungsfunktion $G(z) = 1 - z^{-1/2}$ führt zur Darstellung im kontinuierlichen Frequenzbereich mit:

$$G(\omega) = 1 - e^{-j\omega \cdot T_s/2} = 2j \cdot \sin(\omega \cdot T_s/4) \cdot e^{-j\omega \cdot T_s/4} \quad (\text{IV.35})$$

Die Rauschleistung am Ausgang des Systems wird durch das Quadrat der Übertragungsfunktion bestimmt, die in Bild IV.23 grafisch dargestellt ist:

$$|G(\omega)|^2 = 4 \cdot \sin^2(\omega \cdot T_s/4) \quad (\text{IV.36})$$

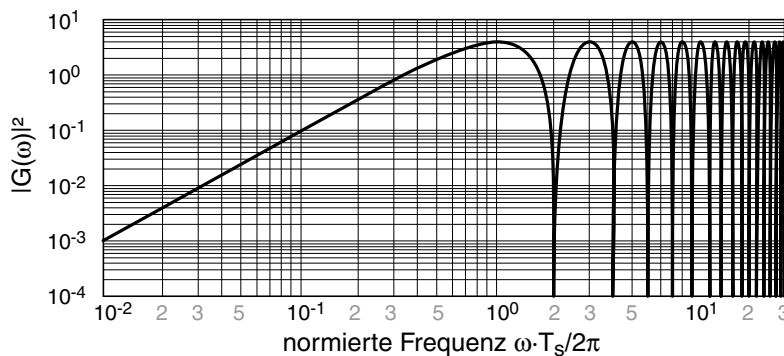


Bild IV.23 Darstellung des Quadrates der Übertragungsfunktion der Korrelierten Doppelabtastung

21. CDS = *Correlated Double Sampling*- Korrelierte Doppelabtastung

Wie zu erkennen ist, erfolgt bei niedrigen Frequenzen eine weitgehende Rauschunterdrückung. Dies betrifft im besonderen das $1/f$ -Rauschen und alle statischen Fehlerterme. Die untere Grenzfrequenz hängt dabei direkt von dem Abtastintervall T_s ab und beträgt:

$$\omega_G = \frac{2\pi}{T_s} \cdot \frac{2}{\pi} \cdot \arcsin\left(\frac{\sqrt[4]{8}}{4}\right) \approx \frac{2\pi}{T_s} \cdot 0.276 \quad (\text{IV.37})$$

Für $\omega > 2\pi/3T_s$ werden die Rauschterme allerdings mit maximal Faktor 2 verstärkt.

In der technischen Implementierung einer Pixelschaltung entspricht dieses Schema dem sogenannten *True CDS*, bei dem der Resetpegel des Pixels und der Wert nach der Integration voneinander subtrahiert werden. In beiden Termen sind die Rauschanteile enthalten, die beispielsweise von den Schwellspannungsvariationen der Reset- und der Auslestufe verursacht werden. Zudem ist das auch kTC -Rauschen des Reset-Transistors als statischer Term enthalten und wird durch die Differenzbildung vollständig entfernt. Nachteil dieser Methode ist jedoch, daß zu jedem Pixel der Initialwert vor der Integration gelesen und über die gesamte Integrationsdauer gespeichert werden muß. Die Speicherung kann entweder im Pixel selbst erfolgen oder in einem Speicher außerhalb der Sensormatrix. In beiden Fällen sind zusätzliche Elemente erforderlich, die das Design vergrößern und weitere Störungen mit sich bringen.

Eine verfahrenstechnische Vereinfachung verwendet daher den Resetpegel nach dem Auslesevorgang des Integrationswertes. Die Differenz eliminiert weiterhin Offsetfehler, jedoch nicht das Resetrauschen. Die Rauschprozesse beim Reset vor der Integration und nach der Integration sind statistisch unabhängig und werden mit der Abtastrate aufgenommen. Folglich geht das Resetrauschen in diesem Fall entsprechend der Übertragungsfunktion des CDS mit dem Faktor 2 in den Meßwert ein. Wegen der einfachen Realisierbarkeit wird dieses Arbeitsschema jedoch häufig verwendet. Eine tiefergehende Analyse des CDS unter Einbeziehung der Abtastung und der Realisierungsmöglichkeiten ist in [Grau03] zu finden.

6 Fehler von Bildaufnahmesystemen

6.1 Allgemeine Betrachtungen

Die Umwandlung eines realen Bildes in ein digitales Abbild geschieht prinzipiell in folgenden Schritten:

- Abbildung auf den Sensor durch ein optisches System von Linsen und Filtern
- Ortsdiskretisierung des Lichtes durch die geometrische Anordnung der einzelnen Pixel auf dem Sensor
- Umwandlung des Lichtes in eine elektrische Größe und deren analoge Vorverarbeitung
- Digitalisierung der ortsdiskreten Intensitäten durch A/D-Wandlung
- Digitale Nachbearbeitung

Die an den einzelnen Verarbeitungsschritten beteiligten Bauteile und Methode weisen verfahrensbedingte Mängel auf, die sich im resultierenden Bild als Störungen widerspiegeln. So kann eine fehlerhafte Optik zu geometrischer Verzerrung, Dispersion, Farbfehlern oder Unschärfe führen. Das reale Bild wird im Allgemeinen durch optische Transformationen (Skalierung, Spiegelung, Translation) auf die aktive Fläche des Sensors abgebildet. Dies geschieht über ein Linsen- oder Spiegelsystem. Da die Linsen bedingt durch Fertigung und Materialeigenschaften nicht perfekt sind, treten bei der Abbildung Fehler auf. Am bekanntesten sind dabei die geome-

trischen Verzerrungen, die bis zu einem gewissen Grad mit Hilfe anderer optischer Linsen ausgeglichen werden können. Verzerrungen können auch durch entsprechende geometrische Transformationen (Skalierung, Translation, Rotation usw.) des digitalisierten Bildes beseitigt werden. Die Korrekturparameter für die einzelnen Geräte lassen sich dabei anhand von geeigneten Testbildern ermitteln.

Übersteigen die Feinheiten und Details des Originalbildes das Auflösungsvermögen der Kamera, so entstehen durch die Diskretisierung örtliche Unterabstufungseffekte (*Aliasing*). Bei praktischen Realisierungen von Pixelzellen ist der lichtempfindliche Teil in der Regel kleiner als die zu einem Bildpunkt gehörende Matrixfläche. Daraus resultiert eine zusätzliche Tiefpaßcharakteristik. Details zwischen den lichtempfindlichen Teilen werden nicht erfaßt. Eine leichte Defokussierung im Linsensystem kann dies zwar teilweise kompensieren, es ergibt sich damit jedoch allgemein ein Verlust an Schärfe. Durch die für unterschiedliche Wellenlängen variierende Brechung können ebenfalls ortsabhängige Unschärfen und Farbfehler durch Dispersion bei der Abbildung von farbigen Objekten entstehen.

Die Umwandlung der unterschiedlichen Lichtintensitäten in elektrische Signale und deren Weiterleitung geschieht bedingt durch die verwendeten Bauteile nicht immer linear. Korrekturmöglichkeiten bestehen in einer geeigneten Übertragungsfunktionen der Ausleseverstärker oder in der Berücksichtigung der Nichtlinearität bei der Analog-Digital-Wandlung. Rauschen und statistische Schwankungen der verwendeten Bauelemente wirken als weitere Störquellen. Die A/D-Wandlung verursacht Quantisierungsrauschen und kann zu zeitlichen Aliasing-Effekten beitragen. Ist es nicht notwendig, ein Bild in Echtzeit aufzunehmen, so können mehrere Bilder des gleichen Objektes aufgenommen und gemittelt werden. Im wesentlichen wird dabei der Signal-Rausch-Abstand erhöht und die Empfindlichkeit verbessert, systematische Fehler können damit jedoch nicht verringert werden. Weitere Fehler sind durch spezifische Realisierungen des Sensors bedingt. So treten z.B. bei der Bildwandlung in der CCD-Kamera Fehler durch die Querschleiffähigkeit der Speicherelemente auf. Die Isolation zu den Nachbarpixeln ist dabei nicht hoch genug, um das Einfließen von Ladungen in die Umgebung zu verhindern. Die sogenannten *Blooming*-Effekte führen zu Kontrastminderungen und Unschärfen.

6.2 Alias-Effekte

Besonders kritisch bei praktischen Bildsensoren sind die Alias-Effekte, die durch die örtliche Diskretisierung entstehen. Der Begriff „*Aliasing*“ stammt dabei aus der Signaltheorie. Nach dem Abtasttheorem von Shannon [Jer77] (bzw. Kotelnikov²²) muß die Abtastfrequenz mindestens doppelt so groß wie die höchste in einer Funktion enthaltene Frequenz sein, um die Funktion fehlerfrei aus den abgetasteten Werten zu rekonstruieren. Ist die Abtastfrequenz kleiner, überlappen sich die Spektren im Frequenzbereich. Das Bild wird dabei unterabgetastet und es treten Alias-Effekte auf. Ein hochfrequentes Signal kann in diesem Fall die gleichen Abtastwerte liefern wie ein niederfrequentes Signal. Man sagt, das niederfrequente Signal ist ein Alias des

22. Das Abtasttheorem wurde zwar 1948 in großer Breite von C. E. Shannon der Fachöffentlichkeit bekannt gemacht, doch bezieht sich Shannon selbst auf frühere Veröffentlichungen. Verfolgt man diese Spuren zurück, so steht heute zweifelsfrei fest, daß das Abtasttheorem in Zusammenhang mit Problemen der Kommunikationstechnik zuerst 1933 von dem russischen Wissenschaftler Vladimir A. Kotelnikov mathematisch exakt formuliert und publiziert wurde. Die Veröffentlichung in einem sowjetischen Konferenzbericht blieb allerdings im Westen bis heute weitgehend unbekannt.

hochfrequenten Signals [Blinn89/1].

Eine Rekonstruktion aus unterabgetasteten Werten prinzipiell nicht möglich. Eine Erhöhung der Abtastrate scheidet in der Praxis jedoch meist aus, da die örtliche Auflösung aus technischen Gründen nach oben begrenzt ist. Eine häufig verwendete Kompensation ist daher die Mittelung über Nachbarbereiche von Kanten, die störende Effekte abmindert (*Anti-Aliasing*). Die bei der Bildaufnahme beobachtbaren Alias-Effekte treten ebenfalls bei der Bildwiedergabe auf Rastergeräten auf. Je nach Art des Alias-Effektes unterscheidet man dort zwischen *Stairstepping* (stufenförmige Intensitätsänderungen entlang einer Linie), *Linebreakup* (Unterbrechungen entlang von Linien), *Crawling* (Erscheinen/Nicht-Erscheinen von dünnen Linien) und *Scintillation* (Größenveränderung bei der Bewegung von Objekten). Die prinzipiellen Möglichkeiten zur Beseitigung von Alias-Effekten bestehen in einer Erhöhung der Abtastrate, Präfilterung oder Postfilterung. Die Abtastrate ist dabei durch das Reproduktionsmedium vorgegeben. Eine Präfilterung scheidet wegen des frühen Informationsverlustes und fehlender Kenntnis über die Geometrie der Szene aus. Daher wird in der Regel eine Postfilterung verwendet [Blinn89/2]. Eine besondere Form von Alias-Effekten tritt bei der Aufnahme von Farbbildern auf, die sogenannten Farbmoirés.

6.3 Farbmoiré

6.3.1 Signaltheoretische Grundlagen

Von den bei Farbbildsensoren mit einem Sensor zur Detektion von Farben verwendeten Farbfiltern ist das Bayer-Mosaik [Bay76] am gebräuchlichsten (Abschnitt 1.2.1). Dabei wird das Bild in Pixel mit den Informationen jeweils zum Rot-, Grün- und Blauanteil unterteilt. Um die vollständige Farbwiedergabe der aufgenommenen Szene zu erhalten, ist es daher notwendig, das Farbmosaik durch Interpolation der fehlenden Werte zu korrigieren. Dieser Prozeß wird auch als *De-mosaicing* bezeichnet. Das besondere Problem dabei läßt sich aus den signaltheoretischen Betrachtungen abgetasteter Funktionen ableiten. Zur Vereinfachung kann dies auch am eindimensionalen Fall veranschaulicht werden.

Betrachtet wird ein vereinfachtes System nach Bild IV.24, bei dem eine Grundfunktion mit zwei versetzten Abtastfunktionen behandelt wird, die beispielsweise den roten und grünen Pixeln einer Bildzeile entsprechen. An einem bestimmten Punkt ist immer nur ein Abtastwert der Funktion bekannt. Der fehlende Wert muß interpoliert werden. Dafür existiert eine Vielzahl von praktischen Methoden, wie z.B. Bilineare Interpolation [NN02], Polynominterpolation, Cok-Interpolation [Cok87], Freeman-Interpolation [Free88], Laroche-Prescott-Interpolation [LP94], Hamilton-Adams-Interpolation [Ham97]. Eine zusammenfassende Übersicht zu Interpolationsverfahren geben beispielsweise Gunturk et al. [GGAS04].

Aus der Darstellung in Bild IV.24 und dem Abtasttheorem ist ersichtlich, daß die maximale Ortsfrequenz der Grundfunktion kleiner als $1/4$ der Frequenz des Rasters $1/\Delta x$ sein muß, um das Abtasttheorem zu erfüllen. Ist dies nicht gegeben, treten Alias-Effekte auf und es ist formal nicht möglich, eine korrekte Interpolation für den fehlenden Abtastwert zu finden. Zur Verringerung von Alias-Fehlern werden bei Digitalkameras daher üblicherweise Tiefpässe als optische Filtervorsätze verwendet. Diese Filter sind bei den meisten Kameras fest vor dem Sensor montiert und dienen der Begrenzung der Ortsfrequenz. In den praktischen Anwendungsfällen

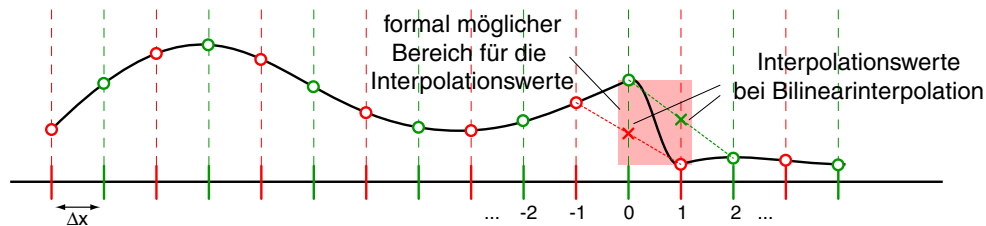


Bild IV.24 Örtlich versetzte Abtastung eines Signals und ein Interpolationsansatz

für Bildaufnahmesysteme sind jedoch gerade die hohen Ortsfrequenzen interessant, da diese als Kanten besonders wichtig für Wahrnehmung und Weiterverarbeitung sind. Eine fehlerhafte Interpolation der Farbwerte führt allerdings zu einem zusätzlichen Farbrauschen im Bild, das aufgrund der höheren Empfindlichkeit der menschlichen Wahrnehmung für Farbunterschiede besonders im roten und blauen Farbbereich verstärkt sichtbar ist (Bild IV.25). Dies wird auch dadurch begünstigt, daß der grüne Anteil im Bayer-Mosaik (Abschnitt 1.2.1) doppelt vorhanden ist. Rot und Blau bestehen in einem derart aufgenommenen Bild zu je 3/4 aus interpolierten Werten. Die Begründung für die Verwendung der doppelten Anzahl grüner Pixel im Vergleich zu Rot und Blau bei Farbfilttern wie dem Bayer-Mosaik ist häufig, daß das menschliche Auge sensitiver für Helligkeiten als für Farben ist bzw. daß sich das Maximum der Hellempfindlichkeit des Auges im grünen Spektralbereich befindet. Mit einer größeren Anzahl grüner Pixel ließe sich daher die Ortsauflösung des Sensors verbessern. Doch diese Schlußfolgerung ist nicht zwingend, da es keinen Grund zu Annahme gibt, daß die Ortsfrequenz in Bildern nur auf die Luminanz beschränkt ist. So ist in [ASH02] und [ASH04] gezeigt, daß ein Sensor exakt die gleiche Ortsauflösung besitzt, wenn im Bayer-Mosaik Rot oder Blau mit Grün vertauscht werden.

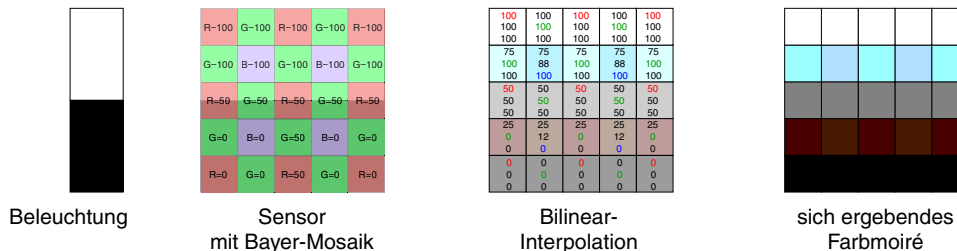


Bild IV.25 Farbmoiré an einem Bayer-Mosaik

Doch auch bei Einhaltung des Abtasttheorems ist eine Farbinterpolation von Fehlern begleitet. Dies ist dadurch begründet, daß die Abtastungen der verschiedenen Farben örtlich versetzt sind und im Bildbereich daher eine Phasenverschiebung von $e^{j\omega \Delta x}$ aufweisen. Die Relevanz der Phasenverschiebung nimmt dabei mit der Ortsfrequenz der abgetasteten Funktion zu. Bei der maximalen Ortsfrequenz von $1/(4 \cdot \Delta x)$ beträgt die Phasenverschiebung $\pi/2$ und die Abtastungswerte nach Bild IV.24 unterscheiden um den vollen Spitzenwert. Bei einer Interpolation kann daher nicht entschieden werden, welcher Wert formal richtig ist, da die Ortsfrequenz unbekannt ist. Folglich ist der örtliche Versatz letztlich die Ursache der als Farbmoiré bekannten Farbsäumen an Bildkanten und nicht die als Tiefpaßfilterung interpretierbare Interpolation, wie in der Literatur vielfach beschrieben [Püg01]. Da der Interpolationswert für jeden Farbkanal auf einer anderen örtlichen Position berechnet wird, ergibt sich folglich ein örtlich verteilter Fehler in den Farbkanälen. Die Art der Farbfehler ist dabei abhängig von Ort und Richtung der lokalen Ortsfrequenz im Bild und macht sich aufgrund der dynamischen Änderung besonders in Bewegtbildern sehr störend bemerkbar.

Zur Verbesserung der Farbinterpolation werden daher von einigen Verfahren explizite Korrelationen zwischen den Farbprimären angenommen. Beispielsweise wird die Hypothese vertreten, daß sich der Quotient von Farbprimären [Kim99] oder deren Differenz [MP02] in natürlichen Bildern nur wenig ändern. Kanten im Bild werden als Farbsprünge beschrieben, die eine andere Interpolation erfordern als innerhalb von homogenen Objekten. Werden dabei in zwei verschiedenen Farbkanälen nahe beieinanderliegende Kanten detektiert, so wird die Interpolation entlang der Kanten unter Berücksichtigung der örtlichen Verschiebung vorgenommen und nicht über die Kanten hinweg. Die *Vector-Median-Interpolation* [GC01] basiert ebenfalls auf diesem Ansatz. Nachteilig ist jedoch der hohe Implementierungsaufwand dieser Verfahren.

6.3.2 Modifikation des Korrekturansatzes

Bei den üblichen Verfahren zum *Demosaicing* wird ein Bild allgemein als farbrichtige Primärfarbdarstellung mit fehlenden Zwischenwerten betrachtet, die es zu interpolieren gilt. Die Sichtweise ist dabei stark wiedergabeorientiert und die Funktion des Sensors wird allgemein ignoriert. Erweitert man die Problemstellung durch Einbeziehung der Sensorfunktion, so ergibt sich das Interpolationsproblem als eine Abtastung des Spektralraumes durch ein zweidimensional angeordnetes Gitter verschiedener Filter. Anstelle der Interpolation der ausgegebenen Abtastwerte kann die Interpolation des originalen auf den Sensor wirkenden Lichtspektrums treten, die anhand der lokalen und der benachbarten Sensorausgaben vorgenommen wird. Daraus lassen sich anschließend die lokal fehlenden Sensorausgaben schätzen. In [ASM03] wird gezeigt, daß Fehler in der Interpolation generell nicht korrigierbar sind, wenn keine Korrelation zwischen den Farbkanälen besteht. In der Praxis sind die spektralen Empfindlichkeiten der verschiedenen Farbkanäle jedoch nicht orthogonal. Damit existiert immer eine Korrelation zwischen den Ausgaben des Sensors, die zur Verbesserung der Interpolation verwendet werden kann. Die Sensorfunktion sei dabei entsprechend Kapitel III, Abschnitt 3.2 linear und integrierend. Die Ausgaben der Abtastungen nach Bild IV.24 ergeben sich demnach zu $R = \int_{\lambda=0}^{\infty} r(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$ und $G = \int_{\lambda=0}^{\infty} g(\lambda) \cdot \varphi(\lambda) \cdot d\lambda$, mit den spektralen Empfindlichkeiten $r(\lambda)$ und $g(\lambda)$ der beiden Farbkanäle sowie der Leistungsdichte des eintreffenden Lichtes $\varphi(\lambda)$. Weiterhin soll $\int_{\lambda=0}^{\infty} g(\lambda) \cdot r(\lambda) \cdot d\lambda \neq 0$ gelten.

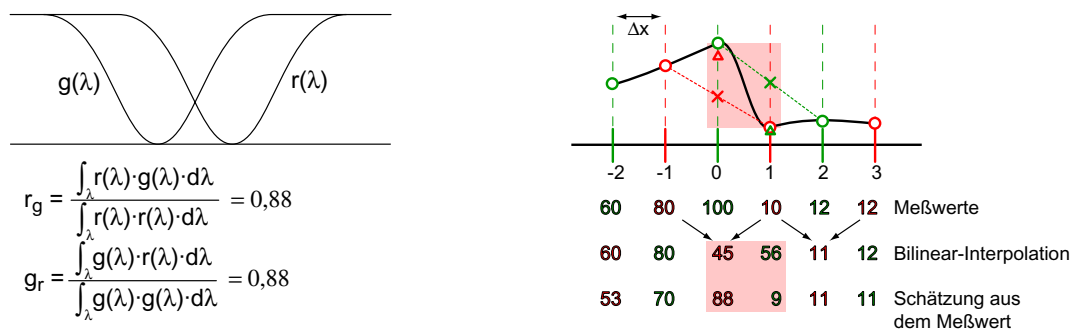


Bild IV.26 Örtlich versetzte Abtastung eines Signals und ein Interpolationsansatz

Der vorgeschlagene Korrekturansatz soll am Beispiel von Bild IV.24 verdeutlicht werden. Betrachtet man die Abtastpunkte für sich, so können diese als einzelne Sensoren mit einem Farbkanal interpretiert werden. Aus den Sensorausgaben kann das originale Eingangsspektrum mit den Rücktransformationsfunktionen nach Kapitel II, Abschnitt 2.2.1 geschätzt werden. Nach Gleichung (II.13) aus ergeben sich diese Rücktransformationsfunktionen $\underline{b}(\lambda)$ allgemein zu

$\underline{b}(\lambda) = \left(\int_{\lambda=0}^{\infty} \underline{a}(\lambda) \cdot \underline{a}(\lambda)^T \cdot d\lambda \right)^{-1} \cdot \underline{a}(\lambda)$, wobei die allgemeine Sensorfunktion $\underline{a}(\lambda)$ im hier betrachteten Beispiel im Falle einer roten Abtastung der skalaren Funktion $r(\lambda)$ und im Falle einer grünen Abtastung der skalaren Funktion $g(\lambda)$ entspricht. Die Rücktransformation erfolgt mit $b_r(\lambda) = \left(\int_{\lambda=0}^{\infty} r(\lambda) \cdot r(\lambda) \cdot d\lambda \right)^{-1} \cdot r(\lambda)$ und $b_g(\lambda) = \left(\int_{\lambda=0}^{\infty} g(\lambda) \cdot g(\lambda) \cdot d\lambda \right)^{-1} \cdot g(\lambda)$. Davon ausgehend können folgende Interpolationen des fehlenden Abtastwertes vorgenommen werden:

- Approximation ohne Beachtung der Nachbarn

Mit der Rücktransfunktionsfunktion kann aus dem gemessenen Wert R bzw. G eine Näherung $\varphi'(\lambda)$ des ursprünglichen Spektrums und daraus eine Approximation des fehlenden Abtastwertes mit $R' = \int_{\lambda=0}^{\infty} r(\lambda) \cdot \varphi'(\lambda) \cdot d\lambda$ und $G' = \int_{\lambda=0}^{\infty} g(\lambda) \cdot \varphi'(\lambda) \cdot d\lambda$ ermittelt werden.

Dabei ergeben sich:

$$R' = \int_{\lambda} r(\lambda) \cdot \left(\left(\int_{\lambda'} g(\lambda') \cdot g(\lambda') \cdot d\lambda' \right)^{-1} \cdot g(\lambda) \cdot G \right) \cdot d\lambda = \frac{\int_{\lambda} r(\lambda) \cdot g(\lambda) \cdot d\lambda}{\int_{\lambda} g(\lambda') \cdot g(\lambda') \cdot d\lambda'} \cdot G = r_g \cdot G$$

$$G' = \int_{\lambda} g(\lambda) \cdot \left(\left(\int_{\lambda'} r(\lambda') \cdot r(\lambda') \cdot d\lambda' \right)^{-1} \cdot r(\lambda) \cdot R \right) \cdot d\lambda = \frac{\int_{\lambda} r(\lambda) \cdot g(\lambda) \cdot d\lambda}{\int_{\lambda} r(\lambda') \cdot r(\lambda') \cdot d\lambda'} \cdot R = g_r \cdot R$$

Ein Beispiel ist in Bild IV.26 zu sehen. Der Approximationsfehler ist bei diesem Korrekturan-satz jedoch stark von den Empfindlichkeiten des Sensors und der spektralen Zusammensetzung des eintreffenden Licht abhängig. Für den Fall identischer Sensitivitäten $r(\lambda)$ und $g(\lambda)$ ist die Korrelation zwischen den Farbkanälen maximal und der Approximationsfehler verschwindet. Der Farbsensor wird in diesem Fall allerdings zu einem Schwarz-Weiß-Sensor. Die Kompen-sation des Farbmosaiktes ist perfekt, jedoch ist keine Farbinformation mehr vorhanden.

Mit einer Vergrößerung der Korrelation der spektralen Empfindlichkeit des Sensors lassen sich demnach zwar die Fehler durch das Farbmoiré verringern, allerdings ergeben sich daraus nega-tive Einflüsse auf die Farbkorrektur. So wurde bereits in Kapitel III, Abschnitt 4.6 gezeigt, daß bei der Berechnung von Farbkoordinaten aus stark korrelierten Sensorausgaben u.a. das Sensor-rauschen verstärkt wird [BH99].

- Interpolation über die Nachbarschaft

Soll im Beispiel nach Bild IV.26 an Position (0) eine Interpolation des roten Abtastwertes aus den Nachbarn vorgenommen werden, so kann auch dies über die Näherung des ursprünglichen Spektrums geschehen. An den Positionen (1) und (-1) berechnet sich $\varphi'_{(k)}(\lambda)$ zu: $\varphi'_{(1)}(\lambda) = b_r(\lambda) \cdot R_{(1)}$ bzw. $\varphi'_{(-1)}(\lambda) = b_r(\lambda) \cdot R_{(-1)}$. An Position (0) ergibt sich die Nähe-rung des Spektrums aus dem gemessenen grünen Abtastwert zu $\varphi'_{(0)}(\lambda) = b_g(\lambda) \cdot G_{(0)}$. Eine Mittelung dieser Spektren kann nun mit $\varphi'_{(0), M}(\lambda) = \sum_k c_k \cdot \varphi'_{(k)}(\lambda)$ erfolgen. Anschließend ergibt sich die Näherung des roten Abtastwertes zu:

$$R'_{(0)} = \int_{\lambda} r(\lambda) \cdot \varphi'_{(0), M}(\lambda) \cdot d\lambda$$

$$= \left[\sum_{i=2k} c_i \cdot \frac{\int_{\lambda} r(\lambda) \cdot g(\lambda) \cdot d\lambda}{\int_{\lambda} g(\lambda) \cdot g(\lambda) \cdot d\lambda} \cdot G_{(i)} + \sum_{j=2k+1} c_j \cdot R_{(j)} \right] \Bigg|_{\forall k}$$

Im einfachsten Fall werden nur die nächsten Nachbarn beachtet. Für die Koeffizienten c_i und c_j können übliche Filterkerne verwendet werden, wobei jedoch zu berücksichtigen ist, wie stark die Korrelation zwischen den Kanälen ist. Ist die Korrelation zwischen den Farbkanälen klein, so kann der lokale Beitrag zur Näherung des fehlenden Abtastwertes nur gering ausfallen. Ist keine Korrelation vorhanden, kann die Interpolation nur aus den Nachbarn gebildet werden. Deren Koeffizienten betragen in diesem Fall bei Beachtung nur der beiden nächsten Nachbarn je $1/2$ und die Gleichung vereinfacht sich zu der bekannten Beziehung der Bilinear-Interpolation mit $R'_{(0)} = (R_{(-1)} + R_{(1)})/2$. Die Bilinear-Interpolation ist demnach ein Spezialfall des vorgeschlagenen Korrekturverfahrens.

Da der Approximationsfehler allgemein auch vom eintreffenden Licht bestimmt sind, können geeignete Koeffizienten aus entsprechenden Testbildern gewonnen werden. Eine weitere Verbesserung der Interpolation ist auf dieser Basis nur durch Analyse des Bildinhaltes möglich, wie beispielsweise bei den Interpolationen nach Kimmel [Kim99] oder Muresan et. al. [MP02], die jedoch nicht auf die Eigenschaften des Sensors eingehen.

- Multikanal-Photodioden

Multikanal-Photodioden ermöglichen durch ihre spektral verschiedenen Kanäle eine prinzipiell bessere Approximation des Eingangsspektrums und somit eine Verbesserung der Interpolation des fehlenden Abtastwertes. Die Vorgehensweise ist mit obigem Beispiel identisch, jedoch treten anstelle der eindimensionalen Sensorsensitivitäten die entsprechend vektorwertigen Funktionen und die Berechnung der Korrektur erfolgt mit den zugehörigen Matrixoperationen.

- Wahrscheinlichkeitstheoretischer Ansatz

Ein weiterer Ansatz für eine Schätzung des fehlenden Abtastwertes kann aus einer wahrscheinlichkeitstheoretischen Überlegung gezogen werden. Für die Wahrscheinlichkeit, daß bei Ausgabe eines Wertes R ein spezifisches Spektrum $\varphi(\lambda)$ am Sensor anliegt, ergibt sich:

$$p(\varphi(\lambda)|R) = \frac{p(R|\varphi(\lambda)) \cdot p(\varphi(\lambda))}{p(R)}$$

Aus den Wahrscheinlichkeiten der Spektren, unter der Bedingung, daß der Wert R gemessen wurde, läßt sich die Wahrscheinlichkeit des fehlenden Abtastwertes G direkt über die Sensorfunktion bestimmen. Da jedoch Wahrscheinlichkeiten über den unendlichdimensionalen Spektrenraum gebildet werden, ist es für praktische Anwendungen sinnvoll, diesen zu beschränken. Dies kann beispielsweise erfolgen, indem nur die für die spezifische Anwendung relevanten Spektren einbezogen werden. Die notwendige Bestimmung der bedingten Wahrscheinlichkeit $p(\varphi(\lambda)|R)$ kann dann beispielsweise mit entsprechenden Testbildern aus dem Anwendungsbereich erfolgen. Der Ansatz deckt sich im wesentlichen mit der Herangehensweise bei klassischen Farbkorrekturverfahren, wobei diese jedoch eine analytische Berechnung von farbkorrigierten Sensorwerten als Ziel haben (siehe auch Kapitel III, Abschnitt 2.1). Die Beschränkung des Ansatzes auf bestimmte Spektren und die relativ aufwendige Realisierung deren Ermittlung begrenzen allerdings die allgemeine Anwendbarkeit dieser Methode.

Kapitel V

Beispielentwurf eines Multikanal-CMOS-Farbsensors

青は藍より出て藍より青し。

1 Grundlagen

1.1 Zielvorgaben

Beim Entwurf eines komplexen Photosensors ist es vorteilhaft, die Konzeption auf Basis der benötigten Funktionalität vorzunehmen. In der vorliegenden Arbeit stand jedoch keine spezielle Anwendung, sondern der Test von Technologie und Entwurfsmethoden im Vordergrund. Daher beruhen die Annahmen über die Funktionalität auf Festlegungen mit einem möglichst breiten und zukunftsorientierten Anwendungsbereich. Die Implementierung von Teststrukturen erfolgt dabei in einer 0,35 µm- und einer 90 nm-Technologie.

Als allgemeine Zielvorgabe wurde ein Farbbildsensor angenommen, der den Leistungsparametern des Auges möglichst nahekommen soll. Der Sensor muß dazu im besonderen die Farben entsprechend der menschlichen Wahrnehmung aufnehmen. Die Realisierung soll mit dem Farbkorrekturschema der spektralen Rekonstruktion erreicht werden. Für eine adäquate Funktion sind daher mehrere Farbkanäle pro Pixel zu implementieren. Die notwendige örtliche Auflösung des Sensors läßt sich jedoch nicht direkt aus der Funktionalität des Auges ableiten. Die Photorezeptoren in der Retina vereinen bei ähnlicher Baugröße wie technische realisierbare Pixelstrukturen weit mehr Funktionalität und sind momentan nicht geeignet reproduzierbar.

Für die allgemeine Anwendung als Bildaufnehmer ist dies jedoch nicht erforderlich, da ein technisches Bildaufnahmesystem immer an ein Bildwiedergabesystem gekoppelt ist. Die notwendigen Parameter für die Aufnahme lassen sich auch von diesem ableiten. Eine Kamera für HDTV(US)²³ benötigt beispielsweise eine maximale Auflösung von 1920x1080 Bildpunkten (ca. 2 MPixel²⁴) bei einer Bildrate von 30 fps²⁵. Aufgrund der Erkenntnisse zum zeitlichen Auflösungsvermögen des Auges ist jedoch eine Erhöhung der Bildrate auf bis zu 100 fps für qualitativ hochwertige Anwendungen durchaus angebracht. Die Pixelgröße ist aus der Forderung zur Auflösung nur mittelbar bestimmt, da noch keine Angaben zur verfügbaren Chipfläche gemacht sind. Nimmt man an, daß dieselbe Optik wie bei handelsüblichen 35 mm Filmkameras²⁶ verwendet werden soll, kann die notwendige Pixelgröße von diesen Abmaßen abgeleitet und mit $35 \text{ mm} / 1920 = 18 \text{ µm}$ festgelegt werden. Die absolute untere Grenze der Pixelgröße ist allgemein durch das Auflösungsvermögen der Optik gegeben und liegt für aktuelle Technologieparameter zwischen 4 und 6 µm [CCGW00]. Eine weitere Verkleinerung der Pixel verbessert die erreichbare Bildqualität nicht. Die maximale Auflösung eines Sensors mit 6 µm Pixelgröße und 24x35 mm aktiver Sensorfläche liegt damit bei 5833x4375 Pixeln (ca. 25 MPixel) und erreicht fast das Auflösungsvermögen der herkömmlichen Fotografie von bis zu 30 MPixel. Die Hersteller von Digitalkamera sind jedoch aus Kostengründen eher den Weg gegangen, die Chipfläche bei vorgegebener Auflösung zu minimieren und anschließend neue Optiken für die sich ergebende Sensorfläche herzustellen. Als geeigneter Parameter für den Entwurf läßt sich daher die kleinste sinnvolle Pixelgröße annehmen. Eine spätere Vergrößerung des

23. *High Definition Television* - Hochauflösendes Fernsehen

24. MPixel - Mega-Pixel, entspricht 2^{20} Pixel

25. fps - *frames per second* - Bilder pro Sekunde

26. Bei 35-mm-Film (Kleinbildfilm) handelt sich um das derzeit am meisten verwendete Filmformat in der Fotografie. Das typische Aufnahmeformat beträgt 24x36mm, es existieren jedoch auch einige Sonderformen, wie das Halbformat mit 18x24mm, das ursprünglich bei 35-mm-Kinofilm verwendet wurde, und das Panoramaformat mit 24x56 mm.

Pixels ist in der Regel problemlos möglich und verbessert die Eigenschaften des Sensors. Die Dimensionierung von Ausleseschaltung, A/D-Wandler, analoger und digitaler Verarbeitung ergeben sich aus dem erforderlichen Datendurchsatz, der direkt von der örtlichen und der zeitlichen Auflösung bestimmt ist. Oben erwähnte HDTV Auflösung bei 30 fps benötigt einen Datentransfer von ca. 62 MPixel/s. Da dies mit einem einzelnen Datenpfad kaum realisierbar ist, sind parallele Auslese- und Wandler-schaltungen erforderlich und bereits in der Konzeption vorzusehen.

1.2 Multikanal-Photosensoren

Wie in Kapitel IV, Abschnitt 3.3.3 gezeigt, ist die spektrale Empfindlichkeit des pn-Überganges einer Photodiode von der Implantationstiefe abhängig. Photodioden in unterschiedlichen Tiefen lassen sich somit zum Design von Farbsensoren verwenden, die zusätzliche Farbkanäle für eine verbesserte Farbkorrektur besitzen.

Die Tiefe der verschiedenen Photodioden ist dabei durch das Dotierungsprofil festgelegt. Die Einflußmöglichkeiten beim eigentlichen Schaltungsentwurf sind auf die Auswahl der entsprechenden Strukturen begrenzt. Allgemein stehen Wannen-Substrat-Dioden und Diffusions-Dioden zur Verfügung. Im hier betrachteten 90 nm-Prozeß liegen Wannendioden ca. 1 µm und Diffusionsdiode ca. 200 nm tief. Vergleichbare Dioden sind in jedem CMOS-Prozeß verfügbar, wobei deren Tiefe jedoch verschieden sein kann. In einem *Tripel-Well*-Prozeß mit isolierten p-Wannen ist noch eine weitere Diode realisierbar [CLSS97], die in der 90 nm-Technologie in einer Tiefe von ca. 700 nm implementiert ist. Diese Dioden können dabei nicht nur lateral, sondern auch vertikal angeordnet werden, wie in Bild V.1a-c illustriert. Dadurch sind pro Pixel mehrere Farbkanäle auf der gleichen optisch aktiven Fläche möglich, die für eine Verbesserung der Farbkorrektur verwendet werden können. Die *Tripel-Well*-Struktur erlaubt zusätzlich noch

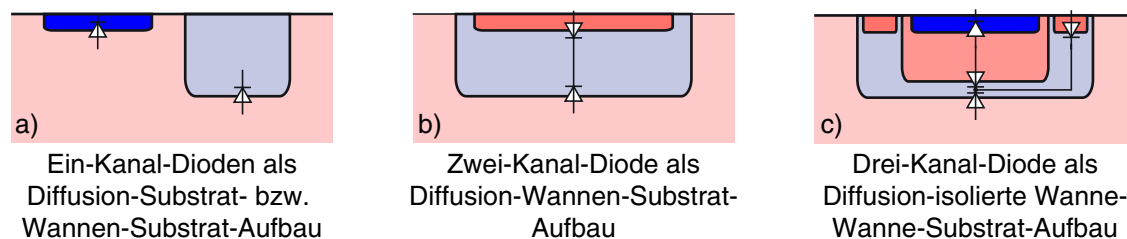


Bild V.1 Mögliche Strukturen für Photodioden im CMOS-Prozeß

eine vierte Diode als nwell/p+ Übergang. Diese besitzt jedoch ähnliche spektrale Eigenschaften wie die p-well/n+ Diode der Basisstruktur, da sie als Diffusionsdiode in der gleichen Tiefe realisiert ist und folglich keinen weiteren farbsensitiven Kanal darstellt. Ähnliche Strukturen sind in [HKS99] charakterisiert.

Vorraussetzung für eine gemeinsame Nutzung dieser Photodioden ist jedoch, daß die verschiedenen Elemente ähnliche absolute Empfindlichkeiten aufweisen, um eine probate Auswertung der verschiedenen Farbkanäle zu gewährleisten. In Kapitel IV, Abschnitt 3.3.3 wurde gezeigt, daß die Photodioden zur Einhaltung dieser Forderung ähnliche Sperrschichtdicken und damit ähnliche Dotierungskonzentrationen aufweisen müssen. Dies kann jedoch in einem Standard-CMOS-Prozeß im allgemeinen nicht erfüllt werden. Die Dotierungskonzentration für Diffusionen übersteigt die Dotierungskonzentration von Substrat und Wannen um ein Vielfaches. Dementsprechend sind auf Basis von Standardprozessen keine qualitativ hochwertigen Multikanal-

Photodioden zu erwarten. Entsprechende Photosensoren auf CMOS-Basis werden folglich Spezialprozesse mit mindestens zwei zusätzlichen Masken und Arbeitsschritten erfordern, bei denen spezielle Implantationen für die Photodioden eingebracht werden müssen.

Zusätzlich ist eine geeignete Wahl der Farbfilter erforderlich, um die Möglichkeiten der Farbkorrektur mittels spektraler Rekonstruktion und die praktischen Gegebenheiten beim Schaltungsentwurf auszunutzen. Die spektralen Eigenschaften der elektronischen Photoelemente sind dabei im wesentlichen durch die Technologie bestimmt und ändern sich bei deren Wechsel. Im herkömmlichen Sensordesign auf Basis des Dreibereichsverfahrens steht die Forderung, daß Farbfilter und Sensor zusammen die Lutherbedingung erfüllen müssen. Somit ist es notwendig, bei jedem Technologiewechsel auch die Farbfilter an die neuen Sensorempfindlichkeiten anzupassen. Prinzipiell können Farbfilter ermittelt werden, die eine exakte Rekonstruktion der gewünschten Zielcharakteristik ermöglichen. Die genaue Nachbildung der CIE-Charakteristik oder einer Linearkombination ist jedoch in der Regel nur mit sehr aufwendigen Methoden möglich [WS82], die allgemein nicht für eine Miniaturisierung geeignet sind. Es stehen somit keine beliebigen Farbfiltercharakteristiken für Mosaikfilter zur Verfügung. Prinzipiell stellen Multikanal-Farbsensoren geringere Anforderungen an die Farbfilter, da sich die höhere Dimension des Sensorraumes zur Farbkorrektur nutzen läßt. Dies kann bereits bei der Filterauswahl berücksichtigt werden. Beispielsweise ist es sinnvoll, daß das Sensorelement möglichst viel Strahlungsenergie empfängt. Dadurch wird eine entsprechende Aussteuerung auch bei geringer Belichtungszeit gewährleistet. Für die Farbfilter ergibt sich, daß in diesem Fall nur ein schmaler Spektralbereich absorbiert werden muß. Im Gegensatz dazu absorbieren herkömmliche Farbfilter Licht über einen weiten Spektralbereich und lassen nur wenig Strahlung bis zum Sensor passieren (siehe auch Kapitel IV, Abschnitt 2.3). Das Problem der Verstärkung des Sensorrauschens bei stark korrelierten Sensorkanälen (vergl. Kapitel III, Abschnitt 4.6) ist dabei unter der Voraussetzung, daß das Photonenrauschen den dominierende Rauschterm des Sensors darstellt, nicht kritisch. Sensoren mit schmalbandigen Farbfiltern werden bei identischen Belichtungszeiten nur wenig ausgesteuert und weisen folglich ebenfalls einen stark verringerten Signal-Rausch-Abstand auf. Eine Korrelation zwischen den Farbkanälen des Sensors ist zudem vorteilhaft für das *Demosaicing* (vergl. Kapitel IV, Abschnitt 6.3.2).

Bei herkömmlichen Farbbildsensoren wird die Genauigkeit der Farbaufnahme allein durch die Farbfiltercharakteristik bestimmt, bei Multikanal-Farbsensoren ist die Zielcharakteristiken jedoch relativ unabhängig von den spektralen Eigenschaften der Farbfilter mit der Korrekturmatrix einstellbar. In den vorgenommenen Analysen war bereits mit acht unabhängigen Sensorkanälen eine angemessene Rekonstruktion des CIE-Normalbeobachters möglich. Auch Technologiewechsel sind bei Multikanal-Farbsensoren unkritisch, da in diesen Fällen nur andere Koeffizienten für die Farbkorrektur benötigt werden. Diese ergeben sich direkt aus der spektralen Empfindlichkeit des neuen Sensorentwurfs. Durch Kombination von verschiedenen Farbfiltern und Photodioden werden verschiedene Realisierungen möglich, die sich hinsichtlich des technischen Aufwandes und der erreichbaren Qualität der Farbaufnahme unterscheiden.

1.3 Auswahl der Schaltungselemente

Als lichtempfindliches Element für die gewählte Zielsetzung ist nur die Photodiode geeignet, da sie als einzige die Realisierung von mehreren spektralen Kanälen pro Pixel und damit Mul-

tikanal-Sensoren erlaubt. Im Standard-CMOS-Prozeß stehen drei verschiedene Dioden zur Verfügung, die jedoch nicht beliebig kombinierbar sind. Zudem gestaltet sich die Berechnung des Quantenwirkungsgrades dieser Strukturen im allgemeinen schwierig, da eine Reihe von notwendigen Parametern, wie beispielsweise Reflexions- und Absorptionseigenschaften der Halbleiterschichten oberhalb der Photodiode im wesentlichen unbekannt und nur durch Messungen am fertigen Schaltkreis zu ermitteln sind. Ebenso sind die Verhältnisse der Dotierung und der lokale Störstellen in der Regel nicht direkt verfügbar. Als Schätzung auf Basis von Messungen an Vorgängertechnologien wird daher angenommen, daß der Quantenwirkungsgrad $\eta_q(\lambda)$ 0.5 nicht übersteigt (Bild V.2).

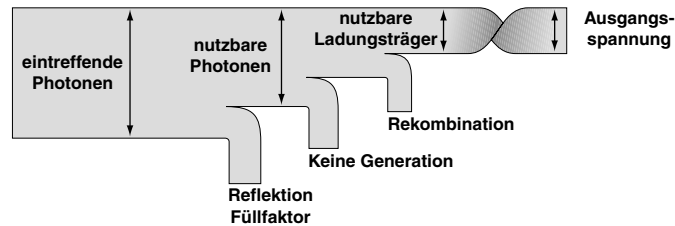


Bild V.2 Modell für die Abschätzung der Quanteneffizienz

Als Pixelschaltung bieten sich Schaltungen mit einfacher Auslese und einem Reset mit MOS-Transistor an, die je wahlweise als NMOS oder PMOS ausgeführt werden können. Dabei ergibt sich der maximal nutzbare Signalhub des Pixels entsprechend folgender Tabelle:

Reset	Spannungsbereich über der Diode	Auslese	Spannungsbereich am Eingang	nutzbarer Spannungshub
NMOS	$0 \dots V_{DD} - V_{t0n}$	NMOS	$V_{t0n} \dots V_{DD}$	$V_{DD} - 2V_{t0n}$
NMOS	$0 \dots V_{DD} - V_{t0n}$	PMOS	$0 \dots V_{DD} - V_{t0p}$	$V_{DD} - \max(V_{t0n}, V_{t0p})$
PMOS	$0 \dots V_{DD}$	NMOS	$V_{t0n} \dots V_{DD}$	$V_{DD} - V_{t0n}$
PMOS	$0 \dots V_{DD}$	PMOS	$0 \dots V_{DD} - V_{t0p}$	$V_{DD} - V_{t0p}$

Es ist zu erkennen, daß die allgemein übliche Schaltungsvariante mit NMOS-Reset und NMOS-Auslese den geringsten nutzbaren Spannungshub besitzt, während die Schaltungen mit mindestens einem PMOS-Transistor nur einen Schwellspannungsverlust aufweisen. Um den Spannungsverlust zu kompensieren, läßt sich zwar mit einer Ladungspumpe die Gate-Spannung der NMOS-Reset-Transistoren erhöhen, dabei sind jedoch die technologischen Grenzen der Durchschlagsfestigkeit zu beachten. Dies ist besonders für moderne CMOS-Prozesse mit sehr dünnem Gate-Oxid und geringer Betriebsspannung von Bedeutung. In der anvisierten 90 nm-Technologie existieren im wesentlichen zwei verschiedene Transistortypen: Digitaltransistoren für eine Betriebsspannung von 1 V (LLD) und Analogtransistoren mit 2.5 V Betriebsspannung (ANA). Eine Erhöhung der Gatespannung ist dabei nicht vorgesehen. Bei Schwellspannungen von ca. 250 mV für Digital- und 470 mV für Analogtransistoren ergeben sich die am Ausgang maximal nutzbaren Spannungshübe zu 750 mV respektive 2030 mV für Schaltungsvarianten mit mindestens einem PMOS-Transistor in der Reset- oder Ausleseschaltung. Da NMOS-Strukturen schon vielfach untersucht wurden, wurden für die bearbeitete Fragestellung in der 90 nm-Tech-

nologie Ausleseschaltungen mit PMOS-Transistoren gewählt. Aufgrund deren Eigenschaft, auch eine Spannung von 0 V über der Photodiode zu übertragen, wird es zudem möglich, die Nichtlinearität des Entladevorgangs der Photodiode genauer zu analysieren. Die Teststrukturen in der 0.35 μm -Technologie bestehen dagegen aus Pixelzellen mit PMOS-Reset und einem traditionellen NMOS-Sourcefolger als Auslesestufe.

Eine Analyse der Transistoren zeigte zudem, daß in der 90 nm-Technologie ein erheblicher Einfluß von Transistorleckströmen auf die Pixelcharakteristik zu erwarten ist. Die Leckströme betreffen dabei im wesentlichen den *Gate-induced-Drain-Leakage* (GIDL) und den *Drain-induced-Drain-Leakage* (DIDL). Beide sind auf Effekte zurückzuführen, die sich aus den geringen Kanallängen der Transistoren und den dafür notwendigen Dotierungsprofilen ergeben. Der GIDL ist ein Tunnelstrom, der bei einem Spannungsunterschied zwischen Gate und Drain durch die Gate-Isolation fließt. Der DIDL dagegen besteht aus Ladungsträgern, die durch die nur wenige Nanometer breite Potentialbarriere zwischen Bulk und Drain tunneln [HYKV00]. Die Auswirkungen der Leckströme betreffen im besonderen den Reset-Transistor und alle Knoten mit Speicherfunktion, die beispielsweise bei einem elektronischen Shutter benötigt werden. Ein Ersatz der Reset-Transistoren durch Dioden ist als Lösung für ein- und auch zweikanalige Photodioden direkt im photoaktiven Bereich implementierbar und ermöglicht zugleich eine verbesserte Flächenausnutzung. Eine Integration des Reset-Transistors in die Photodiode selbst ist ebenfalls möglich, jedoch sind auch hier Verluste durch Leckströme zu erwarten. Der Entwurf von Pixelstrukturen mit elektronischen Shutter kann erst nach einer Messung der Leckströme sinnvoll durchgeführt werden, daher wurden für die Teststrukturen nur einfache Pixelschaltungen implementiert. Bei den für die anvisierte Technologie notwendigen geringen Betriebsspannungen gestaltet sich bereits die Vorverarbeitung der Analogwerte zur Messung der Pixeleigenschaften schwierig. Zur Unterstützung der Analyse wurde daher zusätzlich ein *Rail-to-Rail*-Verstärker entworfen und implementiert, der in Abschnitt 3.2 kurz beschrieben ist.

2 Entwurf von Multikanal-Sensoren in 0.35 μm -Technologie

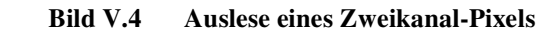
2.1 Implementierung einer Pixelzelle

Der Entwurf in der 0.35 μm -Technologie „AMS035“ dient primär dem Test der Entwurfsmethodik sowie der Gewinnung erster Meßdaten. Aufgrund der Begrenzung von verfügbarer Fläche und Anschlüsse wurde nur ein kleines Testfeld von 16x64 zweikanaligen Pixeln entworfen. Die Pixelzellen bestehen dabei aus einer *Dual-Junction*-Photodiode mit Wannensubstrat- und Wannens-Diffusionsübergang sowie PMOS-Reset und NMOS-Ausleseschaltung (Bild V.3).

Die Funktionsweise des integrierenden Zweikanal-Pixels ist vergleichbar mit der Funktion eines normalen integrierenden Pixels: Zu Beginn der Integrationsphase werden beide Photodioden mittels der Resetschaltung auf die jeweilige Initialspannung rückgesetzt. Dabei ist das Potential auf der Diffusionsdiode gegenüber Masse geringer als das Potential der Wannendiode, so daß beide Photodioden in Sperrichtung vorgespannt sind. Durch den Photostrom wird die Ladung auf den Sperrschichtkapazitäten entsprechend den spektralen Empfindlichkeiten integriert und die Sperrspannungen über Wannendiode und Diffusionsdiode sinken proportional zu Integrationszeit und Bestrahlungsstärke. Dem Schaltungsprinzip entsprechend Bild V.4 werden Potentiale gegenüber Masse nach außen gegeben. Somit ergibt sich die Spannung über der



U_{res2} Select



3.1 Implementierungen von Pixelzellen

118

nem PMOS-Sourcefolger mit PMOS-Schalter (Bild V.5 bis Bild V.7).

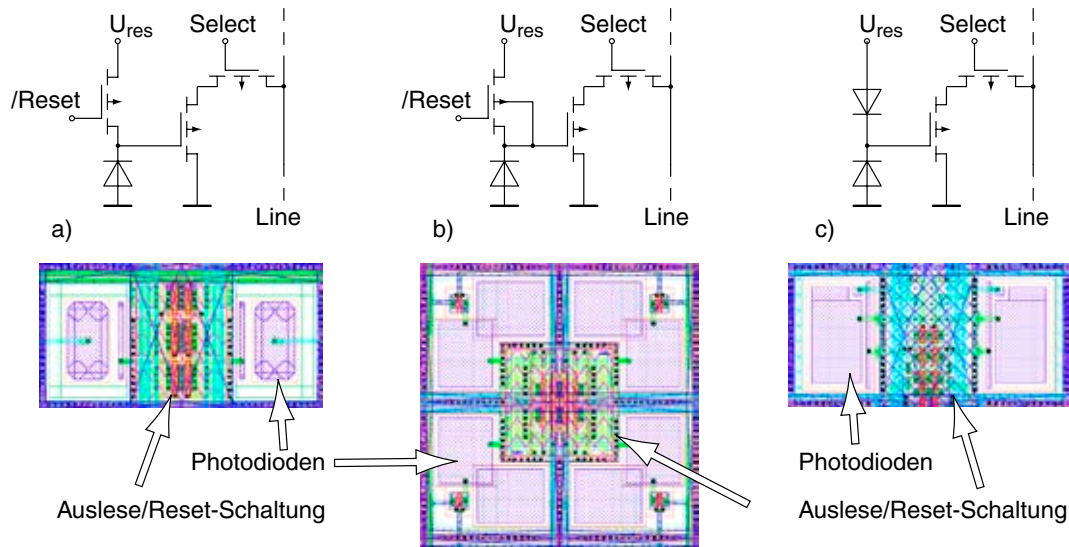


Bild V.5 Schaltungen (oben) und Layoutausschnitte (unten) der realisierten Ein-Kanal-Pixel a) und c) Layouts als Doppel- bzw. b) Vierfachaufbau mit zusammengelegten Auslese-schaltungen von zwei bzw. vier Pixeln

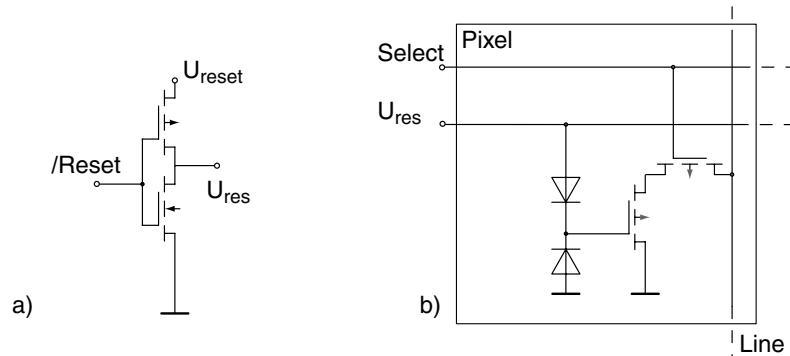


Bild V.6 a) Schaltung zur Erzeugung des Resetsignals für die Pixelschaltung mit Diodenreset bei Diodenreset und b) Pixelstruktur eines Pixels mit Diodenreset

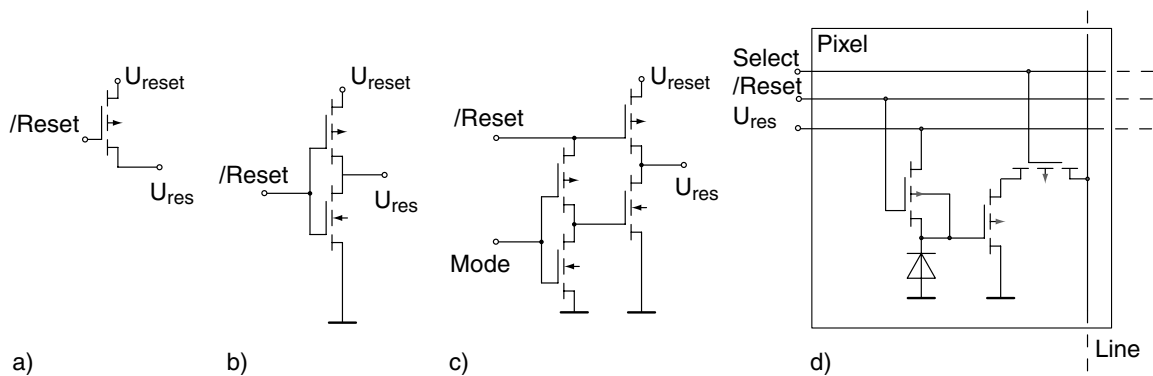


Bild V.7 Zusätzliche Schaltungen zur Erzeugung des Resetsignals für inaktives Reset: a) schwebend, b) auf Masse liegend und c) umschaltbare Variante und d) die Pixelschaltung mit eingebettetem Resettransistor

Die Zwei-Kanal-Photozellen besitzen getrennte Diffusions-Wannen-Substrat-Dioden mit jeweils eigenen Reset- und Auslesestrukturen. Die realisierten Entwürfe beinhalten verschiedene Formen von Diffusionen und *Moat*, um insbes. die spektralen Empfindlichkeiten und die Abhängigkeit des Dunkelstromes von der Diodenform zu untersuchen. Die implementierten Reset-schaltungen sind Transistor-Reset und Dioden-Reset. Die Auslese erfolgt mit PMOS-

Sourcefolger und PMOS-Schalter (Bild V.8). Das Drei-Kanal-Pixel basiert auf dem dreifachen pn-Übergang einer Diffusions-p-Wannen-nWanne-Substrat-Struktur und ist nur in einem *Triple-Well*-Prozeß verfügbar. Das Reset erfolgt mit PMOS-Transistoren und die Auslese mit PMOS-Sourcefolger und PMOS-Schalter (Bild V.9).

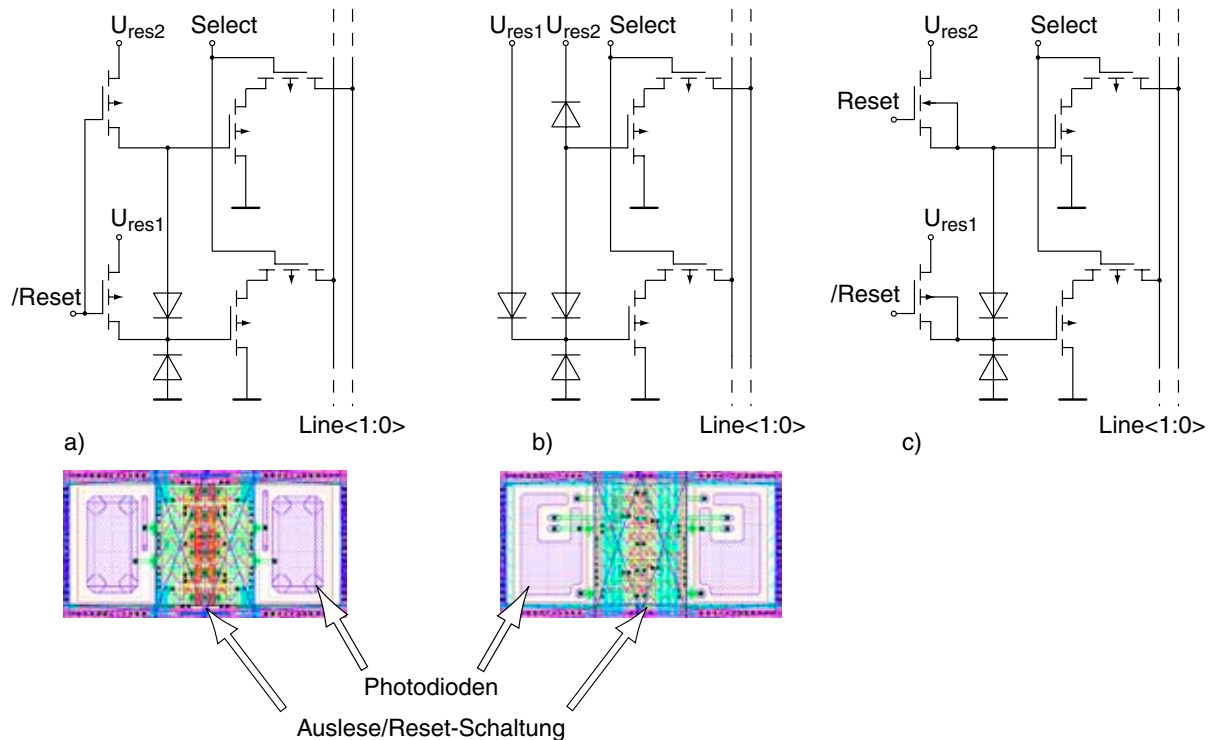


Bild V.8 Schaltungen (oben) und Layoutausschnitte (unten) von Zwei-Kanal-Pixelschaltungen
a) mit klassischem Transistorreset b) Diodenreset c) Reset mit zwei eingebetteten Transistoren (nur a und b auf dem Testfeld realisiert)
Layouts als Doppelaufbau mit zusammengelegten Ausleseschaltungen zweier Pixel

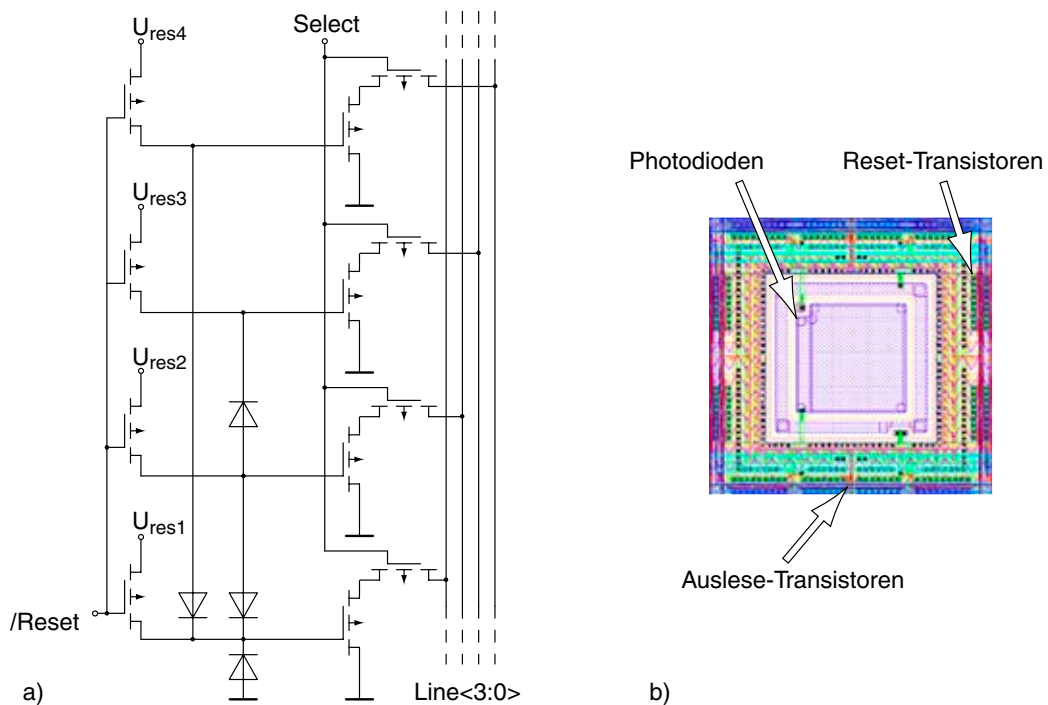


Bild V.9 a) Schaltung und b) Layoutausschnitt des Drei-Kanal-Pixels mit vier Photodioden

Die Pixelzellen wurden jeweils für LLD- und ANA-Transistoren und in verschiedenen Formen entworfen. Es sind jedoch nur wenige Schaltungen mit hauptsächlich ANA-Transistoren implementiert worden, da ANA-Transistoren aufgrund der geringeren Leckströme und des größeren Spannungsbereiches funktionell aussichtsreicher als LLD-Strukturen sind. Für Vergleichszwecke wurde jedoch auch ein Array mit einer Schaltung aus LLD-Transistoren implementiert. Eine Zusammenfassung der implementierten Pixelstrukturen ist in Tabelle V.1 zu finden:

	Transistor -typ	Kanäle	Photodioden	Reset	Auslese
LLD1	LLD	1	nWell-pSubstrat pinned-Photodiode nWell mit Moat	PMOS	PMOS-SF
ANA1	ANA	1	nWell-pSubstrat pinned-Photodiode nWell mit Moat	PMOS	PMOS-SF
ANA2	ANA	2	nWell-pSubstrat & nWell-pDiffusion	Diode	PMOS-SF
ANA3	ANA	2	nWell-pSubstrat & nWell-pDiffusion	PMOS	PMOS-SF
ANA4	ANA	4	nWell-pSubstrat & nWell-pDiffusion & nWell-pWell & pWell-nDiffusion	PMOS	PMOS-SF
ANA21	ANA	1	nWell-pSubstrat pinned-Photodiode nWell mit Moat	Diode	PMOS-SF
ANA22	ANA	1	nWell-pSubstrat pinned-Photodiode nWell mit Moat	eingebetteter PMOS	PMOS-SF

Tabelle V.1 Übersicht der implementierten Pixelschaltungen

Die entsprechenden Layouts der Testfelder sind in Bild V.10 und Bild V.11 zu sehen.



Bild V.10 Testfeld 1 mit Teststrukturen für verschiedene Reset- und Diodenkombinationen

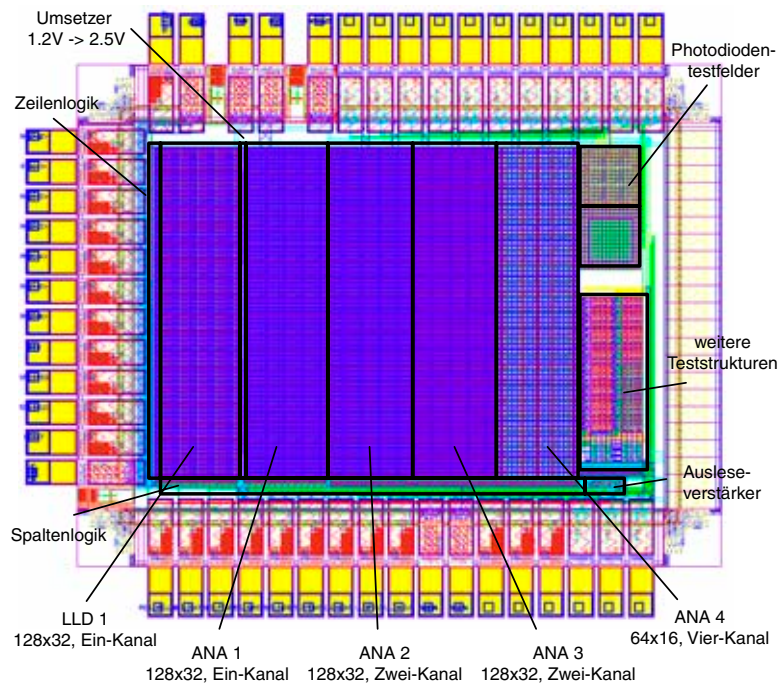


Bild V.11 Testfeld 2 mit Teststrukturen mit LLD- und ANA-Transistoren sowie Einzeldioden

3.2 Operationsverstärker für geringe Betriebsspannungen

Mit der Verringerung der Betriebsspannung wird bei der Signalverarbeitung die Verwendung von sogenannten *Rail-to-Rail*-Verstärkerstrukturen notwendig, die Spannungen über den gesamten Bereich der Versorgungsspannung verarbeiten können. Dazu werden allgemein komplementäre differentielle Paare verwendet [HH96]. Bei einer direkten Zusammenschaltung dieser Paare sind jedoch die geringe Gleichtaktunterdrückung [YES97] nachteilig, da die Transkonduktanz für den Gleichtaktbereich um 100% schwankt (Bild V.12). Für die Funktion als

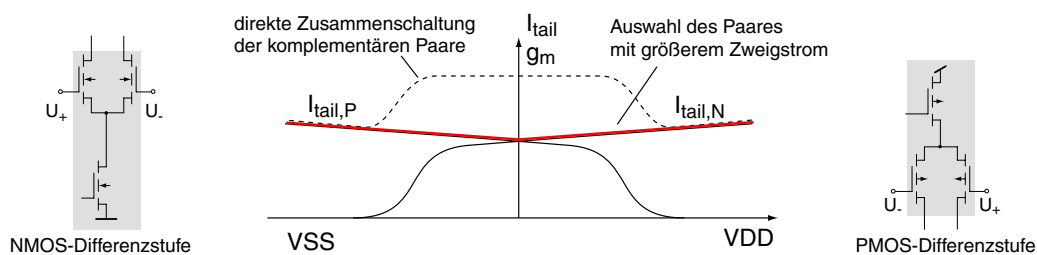


Bild V.12 Eigenschaften komplementärer Differenzstufen

Operationsverstärker ist jedoch eine konstante Transkonduktanz der Eingangsstufe ein wichtiges Kriterium. Ein verbessertes Verhalten kann beispielsweise mit der Kontrolle der Summen der Zweigströme [WMES99]. Eine weitere Möglichkeit besteht in der Auswahl des größeren Zweigstromes. Dabei wird angenommen, daß sich das differentielle Paar mit dem größeren Zweigstrom im gültigen Arbeitsbereich befindet und das jeweils andere Paar ignoriert werden kann. Mit dieser Methode wird die Variation der Transkonduktanz über dem Eingangsspannungsbereich auf ca. 3% reduziert [WMES99].

In Bild V.13 ist die Gesamtschaltung des entworfenen Verstärkers zu sehen und Bild V.14 veranschaulicht die Funktionsweise der Maximum-Ermittlung. Bild V.15 zeigt die simulierte Übertragungskennlinie der Stufe für eine Betriebsspannung von 1.2V.

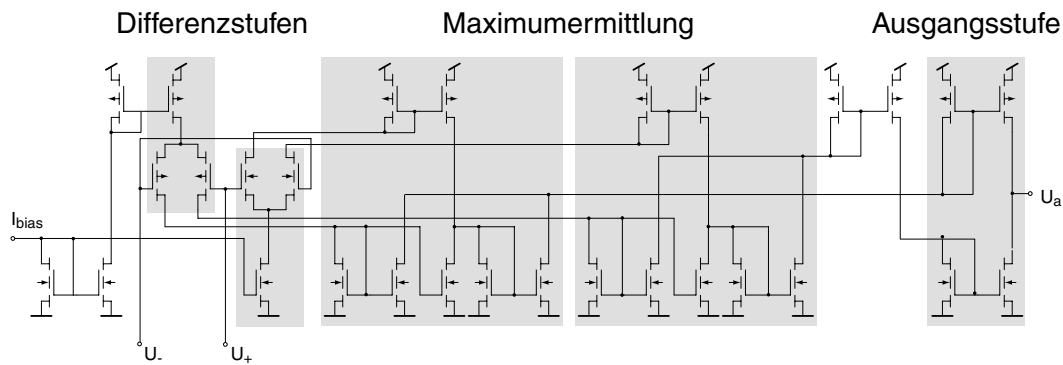


Bild V.13 Gesamtschaltung des Rail-to-Rail-Operationsverstärkers

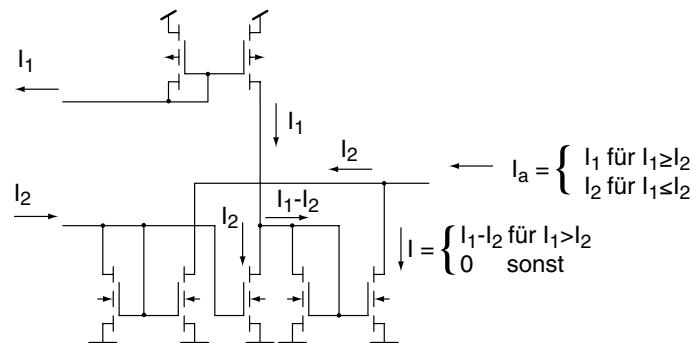


Bild V.14 Maximumermittlung des Rail-to-Rail-Operationsverstärkers

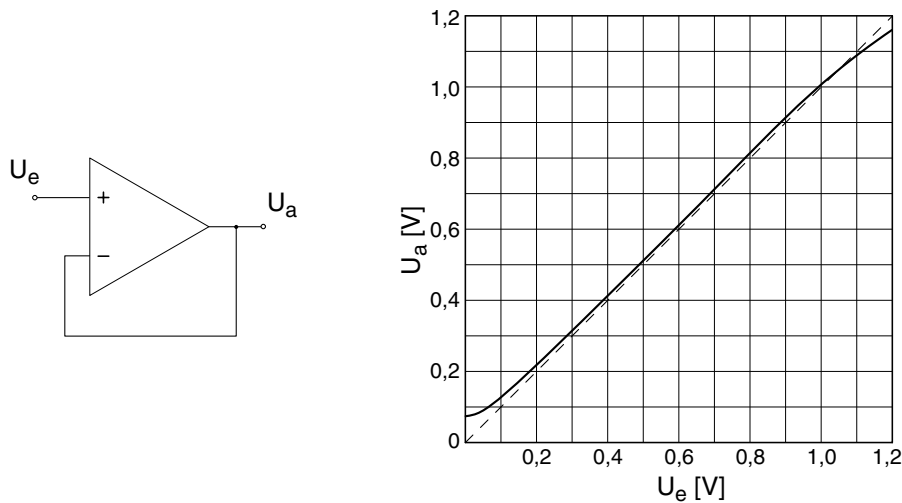


Bild V.15 Kennlinie für Schaltung des Rail-to-Rail-Operationsverstärkers als Buffer

3.3 Funktionelle Erweiterungen der Pixelzellen

Als funktionelle Erweiterungen bieten sich in erster Linie ein elektronische Shutter an (siehe auch Kapitel IV, Abschnitt 4.3.3). Zur Realisierung der Shutter jedoch sind Analogspeicher im Pixel notwendig, die den aufgenommenen Wert über eine spezifische Speicherzeit halten müssen. Der Speicherknoten muß demzufolge ausreichend gut vor Störeinflüssen geschützt sein. Für die betrachtete 90 nm-Technologie kann jedoch davon ausgegangen werden, daß die Leckströme der Transistoren den dominierenden Fehlerterm darstellen, den es zu berücksichtigen gilt. Zur generellen Vermeidung des Einflusses der Transistorleakströme muß der Speicherknoten daher vollständig von allen Drain- und Source-Gebieten getrennt werden. Dies ist beispielsweise mit Dioden in Analogie zu den Resetschaltungen möglich (Bild V.16). Da eine Diode

jedoch nur in einer Richtung leitfähig ist, wird zusätzlich eine zweite Diode zum Rücksetzen des Speicherknotens benötigt. Nachteilig ist dabei neben dem hohen schaltungstechnischen Aufwand, daß der nutzbare Spannungsbereich auf dem Speicherknoten durch die Flußspannungen der Dioden stark eingeschränkt wird. Daher wurde von einer derartigen Implementation abgesehen. Die Besonderheiten der analogen Signalspeicherung bei sub-100 nm-Technologien werden in nachfolgenden Projekten eingehend untersucht.

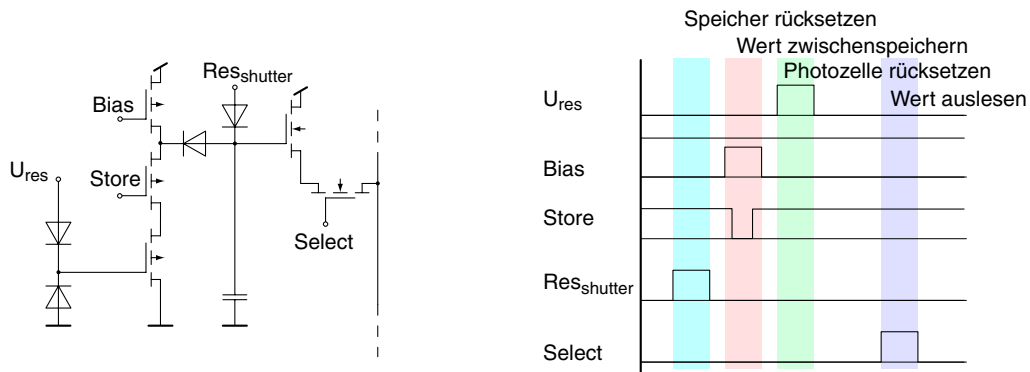


Bild V.16 Elektronischer Shutter mit Dioden

Eine weitere schaltungstechnische Erweiterung stellen Auslesekombinationen dar. Dabei benutzen mehrere Pixelzellen funktionale Teilschaltungen gemeinsam. In der in Bild V.17 gezeigten Schaltung werden beispielsweise der Auslese-Sourcefolger und der Reset-Transistor von vier Pixelzellen genutzt. Der Schaltungsaufwand reduziert sich dabei von zwölf Transistoren für den separaten Aufbau auf sieben Transistoren. Das Auslesen der Zellen kann in diesem Fall allerdings nur sequentiell erfolgen. Zusätzlich ermöglicht diese Schaltung, den Mittelwert über die gemeinsam auslesbaren Pixel zu bilden. Dazu werden die Select-Transistoren gleichzeitig geschlossen. Dies kann beispielsweise bei schlechten Lichtverhältnissen zur Reduktion des Rauschens dienen, wobei sich jedoch gleichzeitig die optische Auflösung verringert.

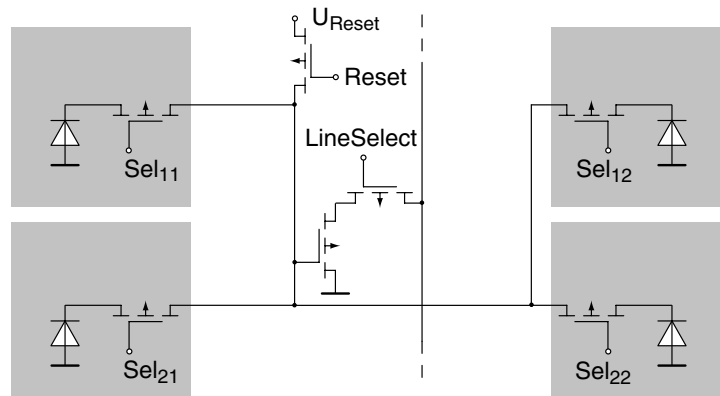


Bild V.17 Vier-Pixel-Kombination mit gemeinsamer Auslese- und Rücksetzschaltung

3.4 Konzeption einer Auslese/Korrekturschaltung

Zur Realisierung eines Kamerasystems sind neben den photoempfindlichen Sensoren weitere Verarbeitungseinheiten notwendig. Moderne CMOS-Prozesse besitzen jedoch relativ schlechte Parameter bezüglich analoger Signalverarbeitung. Höhere Leckströme, stärkerer Mismatch und ein eingeschränkter Betriebsspannungsbereich im Vergleich zu herkömmlichen Prozessen wirken sich negativ auf den ausgangsseitigen Dynamikumfang aus. Mit schaltungstechnischen und

algorithmischen Maßnahmen lassen sich dabei jedoch einige Verbesserungen erzielen. Als Beispiel soll das Konzept einer Mixed-Signal-Implementation der Farbkorrektur basierend auf der spektralen Rekonstruktion dienen. Diese läßt sich als Teil der Ausleseschaltung zusammen mit A/D-Wandler, Defektpixelkorrektur sowie Interpolations- und Filteralgorithmen zur Unterdrückung von (Farb-)Moiré realisieren.

Die Korrektur mittels spektraler Rekonstruktion verwendet die Linearkombination der verschiedenen Farbkanäle des Sensors. Da die Farbkanäle bei Verwendung einer lateralen Anordnung von Absorptionsfiltern ebenfalls räumlich verteilt sind, kann diese Umrechnung als Faltung des Sensorbildes mit einer Faltungsmaske interpretiert werden. Eine einfache Defektpixelkorrektur kann ebenfalls eine Faltungsoperation interpretiert werden. Beispielsweise kann die Ausgabe eines Sensorelements im Falle eines Defektes durch den Durchschnitt seiner nächsten Nachbarn ersetzt werden. Diese Funktion muß dabei für jedes Sensorelement individuell programmierbar sein. Filteralgorithmen zur Moiré-Unterdrückung, wie z.B. die Bilinearfilterung sind ebenfalls räumliche Faltungen. Zur Realisierung aller Funktionen ist somit nur eine einzige Verarbeitungseinheit notwendig. Die Operationen können zudem gleichzeitig ausgeführt werden, da die Faltungsmasken unabhängig vom Sensorbild extern zusammengefaßt werden können (Bild V.18).

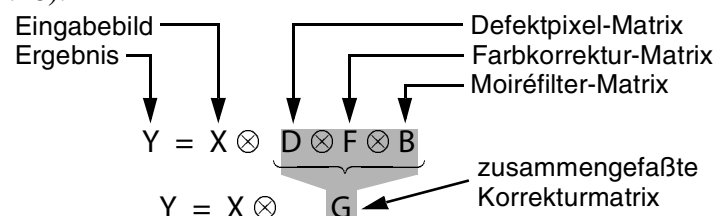


Bild V.18 Zusammenfassung der Faltungsoperationen

In [SGGS01] ist eine Struktur vorgestellt, die mittels Mixed-Signal-Vorverarbeitung die Faltung eines Sensorbildes mit programmierbaren Filtermasken ermöglicht. Die Berechnung des Ergebnisses erfolgt dabei bitsequentiell (Bild V.19).

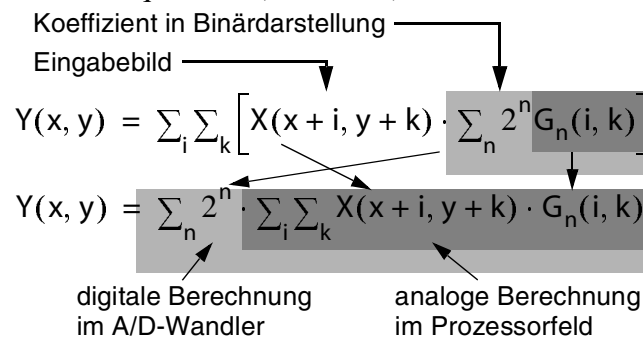


Bild V.19 Aufteilung der Faltungsoperation

Die Ausgaben der verschiedenen Sensorelemente werden zunächst mit einem CDS-Ausleseverfahren von Offsetfehlern befreit und zwischengespeichert. In analogen Prozessorfeldern [SGGS00] erfolgt die bitweise Multiplikation der analogen Sensordaten mit digitalen Koeffizienten und die Summierung der Produkte. Bei der Wandlung mit einem speziellen Pipelined Analog-Digital-Konverter [SGSG01] erfolgt die digitale Akkumulation der Partialsummen. Die Ausgabe besteht dabei mindestens aus den drei Farbkoordinaten des Zielfarbraumes. Im Falle der vorgestellten Farbsensoren mit getrenntem Infrarotkanal sind vier Berechnungen notwendig. Diese können seriell nacheinander oder parallel in je einem eigenen Prozessorfeld berech-

net werden. Die Architektur nach Bild V.20 ist skalierbar und ermöglicht die Ermittlung korrigierter Farbkoordinaten, eine programmierbare Defektpixelkorrektur, eine Moiré-Unterdrückung und sowie weitere Filteroperationen. Die Struktur ist kompakt und erfordert vergleichsweise geringe Anforderungen an Komplexität, Flächenbedarf und Verlustleistung.

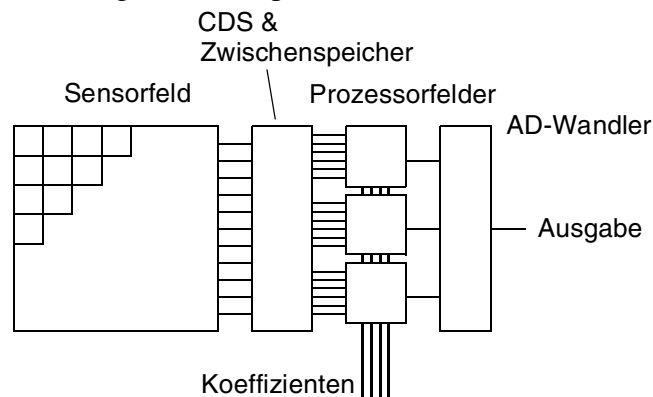


Bild V.20 Architektur eines Sensorsystems mit Korrekturschaltungen

Im Rahmen dieser Arbeit wurden Ausleseschaltungen jedoch nur konzipiert. Die Validierung des Auslese- und Korrekturprinzips erfolgte durch Systemsimulation und die Implementierung ist für folgende Projekte vorgesehen. Aufgrund der hohen Komplexität von Systemsimulationen unter Verwendung herkömmlicher Netzwerkanalyseprogramme, wurde eine eigene Simulationsumgebung entwickelt, deren funktionale Grundlagen in [HGG00/1] und [HGG00/2] beschrieben sind. Weitere Anwendungen des Simulators finden sich in [Hen04/1] und [GGS01].

4 Meßergebnisse

Zur Zeit liegen nur Meßergebnisse des Testfeldes in AMS035-Technologie vor. Dieses Testfeld kann zwar prinzipiell zur Bildaufnahme verwendet werden, ist dazu jedoch aufgrund der geringen Größe von 16x64 Pixeln nur für einen beschränkten Einsatzbereich geeignet. Es ist für grundlegende Untersuchungen bezüglich der spektralen Empfindlichkeit von Photodioden, statistischer Bauelementeschwankungen sowie Randeffekte beim Matrixaufbau vorgesehen. Bild V.21 zeigt eine Beispielaufnahme des Testfeldes.



Bild V.21 Beispielaufnahme des Testfeldes

Die spektrale Empfindlichkeit der Photodioden wurde mit Hilfe eines Monochromators aufgenommen und ist in Bild V.22 gezeigt. Gemessen wurden die Änderungen der Spannungen ΔU_{Well} und ΔU_{D} der Dioden bezüglich Massepotential, die sich bei Bestrahlung mit Licht einer bestimmten Wellenlänge und Bestrahlungsstärke über einer Integrationszeit ergeben. Die Sperrspannung über der Diffusionsdiode ergibt sich dabei zu $\Delta U_{\text{Diff}} = \Delta U_{\text{Well}} - \Delta U_{\text{D}}$ entsprechend Bild V.4. Da zur Erzeugung der monochromatischen Bestrahlung ein Temperaturstrahler verwendet wurde, ist es notwendig, dessen spektrale Charakteristik in einem zusätzlichen Kalibrierschritt aus den Meßwerten herauszurechnen. Aufgrund der dadurch verstärkten Meßfehler werden besonders bei geringen Wellenlängen Diskrepanzen in der Darstellung verursacht.

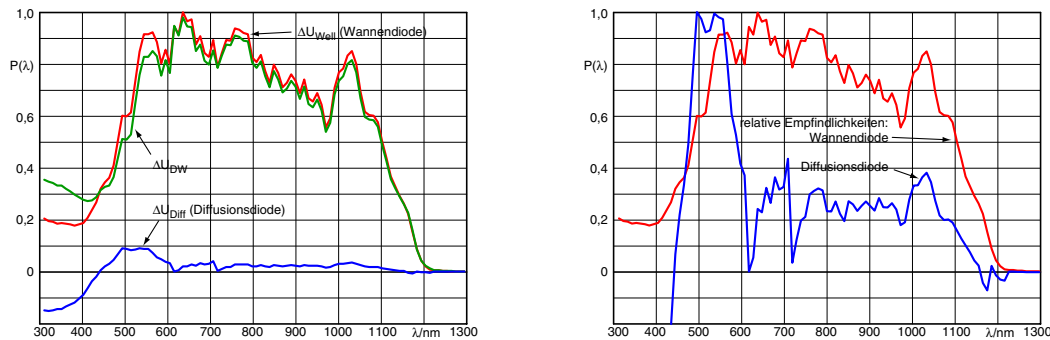


Bild V.22 Absolute und relative spektrale Empfindlichkeit der Photodioden (normiert)

Die Messungen bestätigen jedoch prinzipiell das erwartete Verhalten: Die Diffusionsdiode ist besonders für kurze Wellenlängen empfindlich, die tiefer liegende Wannendiode ist im Vergleich dazu empfindlicher für große Wellenlängen. Die Wannendiode zeichnet sich zudem durch eine wesentlich höhere absolute Empfindlichkeit aus. Dies ist auf den Unterschied der Dotierungskonzentrationen zurückzuführen. Die Diffusionsdiode ist stärker dotiert und weist somit eine kleinere Raumladungszone und höhere Sperrschichtkapazität auf (vergl. Kapitel IV, Abschnitt 3.3.3). Aufgrund der stark unterschiedlichen absoluten Empfindlichkeiten ist folglich eine gemeinsame Verwendung dieser Dioden im vorgestellten Farbkorrekturschema nicht aussichtsreich. Bild V.23 verdeutlicht dies am Beispiel einer Farbaufnahme mit dem Testfeld unter Verwendung von drei Farbkäneln mit externen Filtern in Rot, Grün und Blau sowie einem zusätzlichen cyanfarbenen Infrarotfilter.



Bild V.23 Original (links) und Farbbildaufnahme des Testfeldes (rechts)

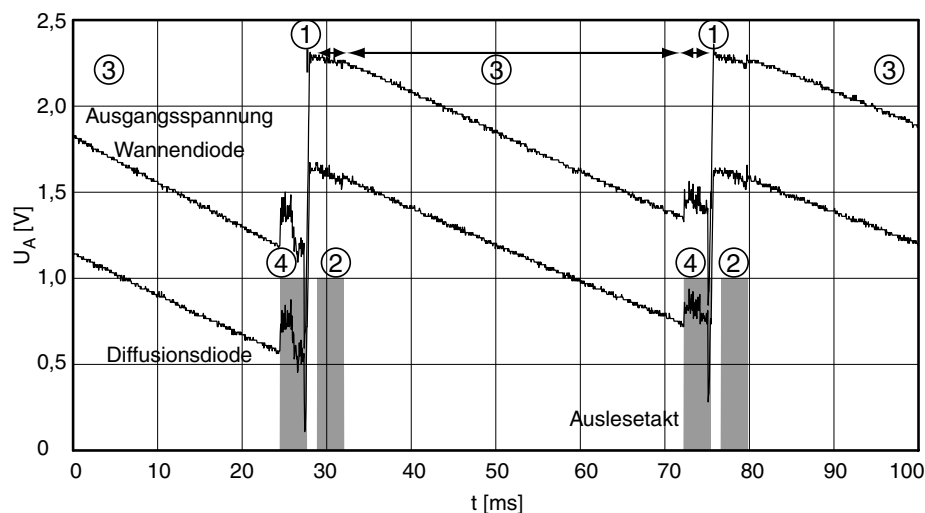


Bild V.24 Zeitverlauf eines Auslese- und Integrationsvorganges

In Bild V.24 ist die Messung eines Auslese- und Integrationsvorganges ausschnittsweise für eine Bildzeile mit *True-CDS* (CDS - *Correlated Double Sampling*) im Einzelzeilenmodus dargestellt. Dabei erfolgt zu Beginn ein Reset der Bildzeile (Phase 1) und ein sequentielles Auslesen des Resetzustandes der 16 Spalten (Phase 2). Nach der Integrationszeit (Phase 3) wird der Integrationswert der Spalten ausgelesen (Phase 4). Anschließend wird der Vorgang mit der nächsten Zeile wiederholt. Die Aufnahme von Reset- und Integrationswert erlaubt die Eliminierung

von Offsetfehlern und Resetrauschen mittels Differenzbildung (vergl. Kapitel IV, Abschnitt 5.4.1). Für die Untersuchung von Randeffekten wurde die Testmatrix in unterschiedlicher Entfernung zu Randstrukturen im Gesamtlayout platziert. Dadurch wird es möglich, die Auswirkung von Inhomogenitäten auf eine Pixelmatrix und auf die Funktionalität der Photosensoren zu analysieren. In Bild V.25 wurde das aufgenommene Originalbild (vergl. Bild V.21) in ein stilisiertes Layout des Schaltkreises eingefügt. Die Darstellung zeigt, daß im ungünstigsten Fall 5 Pixelzellen am Rand beeinflusst und daher nicht zur Bildaufnahme geeignet sind. Schaltungs- teile, die direkt an die Sensormatrix angrenzen, verringern die Randeffekte, können sie jedoch nicht gänzlich vermeiden. Dies ist nur durch einen Ring von mehreren Pixelzellen möglich, die nicht zur eigentlichen Pixelmatrix gehören, jedoch ein homogenes Sensorfeld sicherstellen.

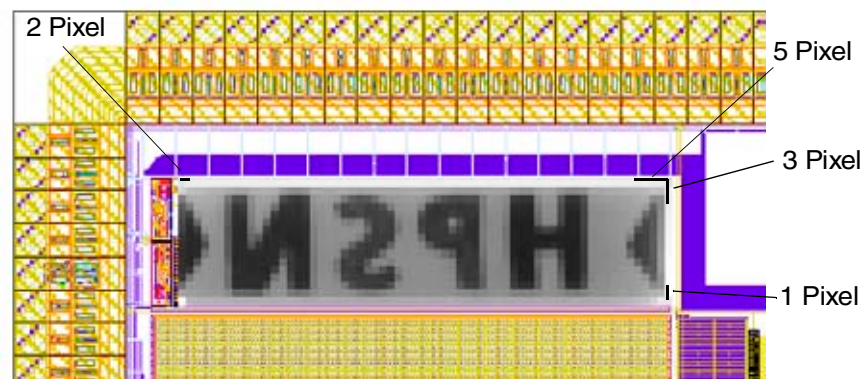


Bild V.25 Größe der Randeffekte an den verschiedenen Seiten der Matrix
(Layout seitenrichtig, eingeblendetes Bild durch die optische Abbildung spiegelverkehrt)

Zusammenfassung und Ausblick

Ein wichtiges Kriterium von Systemen zur Farbbildaufnahme ist die Fähigkeit, Farbbilder im gesamten menschlichen Wahrnehmungsbereich korrekt aufzuzeichnen. Die aktuell verfügbaren Systeme können diese Forderung jedoch im allgemeinen nicht erfüllen. Bei der Betrachtung von herkömmlichen Farbbildsensoren läßt sich schlußfolgern, daß eine Farbmessung mit drei Farbkkanälen entsprechend dem Dreibereichsverfahren unter den gegebenen technischen Voraussetzungen nur unzureichend möglich ist. Farbbildsensoren sind aufgrund der technischen Sachzwänge folglich auch in Zukunft von einer zusätzlichen Farbkorrektur abhängig. Der Bedarf an verbesserten Aufnahmesystemen wird in Zukunft insbesondere mit der Einführung neuer Wiedergabemedien, wie z.B. HDTV, stark ansteigen.

Aus der biologischen Funktionalität des Auges wurde abgeleitet, daß die menschliche Farbwahrnehmung nur von den spektralen Absorptionseigenschaften der Sehpigmente abhängig ist und nicht durch die Weiterverarbeitung der Reize bis zum Gehirn beeinflußt wird. Die spektrale Empfindlichkeit der Sehpigmente weist allerdings genetisch bedingte, individuelle Unterschiede auf. Die Farbwahrnehmung ist folglich ein subjektiver Vorgang. Durch die Standardisierung nach CIE / DIN5033 ist jedoch eine relativ einheitliche Beschreibung der durchschnittlichen menschlichen Farbwahrnehmung möglich.

Das aus der Standardisierung abgeleitete Modell der metameren Farbreproduktion ist primär für die Anwendung auf dem Gebiet der Farbwiedergabe ausgelegt. Eine allgemeine Beschreibung der Farbaufnahme ist damit nur eingeschränkt möglich. In dieser Arbeit wurde daher eine Erweiterung des Farbwahrnehmungsmodells entwickelt. Als Grundlage wird die verallgemeinerte Betrachtung der Farbwahrnehmung als eine spezielle Integraltransformation verwendet. Die spektralen Empfindlichkeiten der betrachteten Photorezeptoren bilden dabei die Basisfunktionen der Transformation. Es konnte gezeigt werden, daß die Wahrnehmung von zwei unterschiedlichen Photorezeptoren äquivalent ist, wenn deren Basisfunktionen linear abhängig sind.

Dieses Farbwahrnehmungsmodell dient als theoretische Grundlage der entwickelten Methode der spektralen Rekonstruktion zur Farbkorrektur. Es wurde gezeigt, daß eine Farbkorrektur von fehlerbehafteten Sensorausgaben allgemein nur durch deren Linearkombination möglich ist. Nichtlineare Korrekturansätze können eine Fehlerreduktion nur für ein spezifisches Eingangsspektrum garantieren. Die vorgestellte Farbkorrektur bietet dagegen eine vom Eingangslicht unabhängige Korrektur. Fehlerbehaftete Sensorausgaben werden dabei in Ausgaben eines idealen Sensors umgerechnet, dessen Parameter frei gewählt werden können. Bei den in dieser Arbeit aufgeführten Beispielen ist als Idealsensor zur besseren Vergleichbarkeit jeweils der CIE-Normalbeobachter verwendet worden.

Die Methode der spektralen Rekonstruktion gewährleistet eine optimale Farbkorrektur. Als Bewertungskriterium dient der quadratische Fehler zwischen der erreichten und der vorgegebenen spektralen Charakteristik. Im Gegensatz zu dem in der Farbmessung üblicherweise verwendeten Farbabstand bietet der Quadratfehler eine Bewertung, die unabhängig von der spektralen

-

Zusammensetzung des zu messenden Lichtspektrums ist. Der Vergleich der spektralen Empfindlichkeit des farbkorrigierten Systems mit der Zielvorgabe erlaubt weiterhin detaillierte Aussagen zu Art und Größe von unkorrigierbaren Fehlern, die durch die eingeschränkten technischen Möglichkeiten bei der Herstellung von optischen Filtern bedingt sind.

Die Methode der spektralen Rekonstruktion eignet sich besonders aufgrund ihrer Flexibilität als allgemeines Farbkorrekturverfahren. Die Korrektur kann mit einer beliebigen Anzahl von Eingangs- und Ausgangskanälen vorgenommen werden. Die Verwendung von mehreren spektral unabhängigen Eingangswerten ermöglicht darüberhinaus eine Verbesserung der Korrektur. Bei Farbbildsensoren eröffnet das Verfahren somit die Möglichkeit, durch Verwendung zusätzlicher Farbkanäle eine verbesserte Farbaufnahme im Vergleich zum herkömmlichen Sensordesign mit drei Farbkanälen und eine Annäherung an spektrale Farbmeßverfahren zu erreichen.

Das allgemeine Modell der Farbwahrnehmung ermöglicht des weiteren eine einheitliche Beschreibung der beiden standardisierten Farbmeßverfahren, Spektralmeßverfahren und Dreibeereichsverfahren. Ebenfalls läßt sich die nach DIN5033 empirisch vorzunehmende Parameteranpassung für Teilmeßfilter mit der Methode der spektralen Rekonstruktion durch ein analytisches Optimierungsverfahren ersetzen.

Um die Möglichkeiten der Farbkorrekturmethode zu untersuchen, wurden in einer Reihe von Implementierung Teststrukturen neuer Multikanal-Sensoren entworfen. Dabei wurden sowohl neuartige Schaltungen entwickelt als auch die Entwurfsmethoden und Besonderheiten der Implementierung analoger CMOS-Schaltungstechnik in sub-100nm-Technologien erprobt.

Eine besondere Herausforderung der neuen CMOS-Prozesse besteht in den stark angestiegenen Leckströmen. SC- und SI-Schaltungstechniken, die auf der Speicherung von analogen Werten beruhen, sind von den verschlechterten Bauelementeparametern besonders betroffen.

Die Schaltungsentwürfe konzentrieren sich daher bei integrierenden Pixelzellen mit Multikanal-Photodioden auf die Kompensation von möglichen Störungen an der Integrationskapazität. Ein im Rahmen dieser Arbeit entwickeltes neues Schaltungskonzept zur Einstellung des Anfangswertes auf der Integrationskapazität wurde patentiert.

Zukünftige Herausforderungen auf dem Gebiet von Farbbildsensoren in modernen CMOS-Technologien liegen in der technologischen Realisierung von geeigneten Dotierungsprofilen für Multikanal-Photodioden und in der technische Implementierung von Ausleseverfahren, die Gegenstand aktueller Forschungen sind. Ein weiteres Problem stellt die Erweiterung des relativ eingeschränkten optischen Dynamikbereiches aktueller Kamerasysteme dar. Weiterhin werden die Möglichkeiten untersucht, die sich aus der Kombination der Multikanal-Farbbildaufnahme und neuen Wiedergabegeräten mit mehr als drei Primärfarben ergeben.

Auf dem Gebiet der Farbinterpolation sind weitere theoretische Untersuchungen notwendig, um offene Fragen zu klären. Erste Ansätze zu deren Lösung konnten bereits im Rahmen dieser Arbeit vorgestellt werden.

Literaturverzeichnis

- [AKP87] P.K. Ahnelt, H. Kolb, R.Pflug, Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina, *J Comp Neurol*, 255(1):18–34, 1987.
- [ASH02] D. Alleysson, S. Süsstrunk, J. Héroult, Color demosaicing by estimating luminance and opponent chromatic signals in the Fourier domain, *Proc. IS&T/SID 10th Color Imaging Conference*, pp. 331–336, 2002.
- [ASH04] D. Alleysson, S. Süsstrunk, J. Héroult, Linear demosaicing inspired by the human visual system. Accepted for publication in *IEEE Transactions on Image Processing*, 2004.
- [ASM03] D. Alleysson, S. Süsstrunk, J. Marguier, Influence of Spectral Sensitivity Functions on color demosaicing, *Proc. IS&T/SID 11th Color Imaging Conference*, pp. 351–357, Scottsdale, AZ, November 2003.
- [Bay76] B.E. Bayer, Color imaging array, US Patent 3 971 065, 1976.
- [BH99] R.L. Baer and J. Holm, A model for calculating the potential ISO speeds of digital still cameras based upon CCD Characteristics, *Proc. PICS 1999: Image Processing, Image Quality, Image Capture, Systems Conference*, 1999.
- [BK69] B. Berlin, P. Kay, *Basic Color Terms*, Berkeley, Calif.: University of Calif. Press, 1969.
- [Blinn89/1] J.F. Blinn, Dirty Pixels, *IEEE Computer Graphics and Applications*, pp.100–105, Juli 1989.
- [Blinn89/2] J.F. Blinn, Return of the Jaggy, *IEEE Computer Graphics and Applications*, pp.82–89, März 1989.
- [Böhm97] M. Böhm, *Mikroelektronik Teil 35 CCD-Bildsensoren*, Universität-Gesamthochschule Siegen, 1997.
- [Boll1877] F.C. Boll, *Zur Physiologie des Sehens und der Farbempfindung*, Kgl. Akad. der Wissensch. zu Berlin, 1877.
- [Bös87] K.D. Bösing, Über Aspekte der visuellen Wahrnehmung und der optischen Eigenschaft idealer Oberflächen, *Technischer Bericht 87/12*, Technische Universität Berlin, Institut für Technische Informatik, Computer Graphics & Computer Vision, 1987.
- [Bron87a] I.N. Bronstein, K.A. Semendjajew, *Taschenbuch der Mathematik*, 23. Auflage, pp.749–754, Verlag Nauka, Moskau, 1987.
- [Bron87b] I.N. Bronstein, K.A. Semendjajew, *Taschenbuch der Mathematik*, Erweiterungsband, pp.111–116, Verlag Nauka, Moskau, 1987.
- [Camp93] C.v. Campenhausen, *Die Sinne des Menschen*, G. Thieme Verlag, Stuttgart, 2. Edition, 1993.
- [CCGW00] T. Chen, P. Catrysse, A. El Gamal, B. Wandell, How Small Should Pixel Size Be?, In *Proceedings of SPIE Electronic Imaging '2000*, Volume 3965, Januar 2000.
- [CIE86] Publication CIE No 15.2, *Colorimetry*, Second Edition, Vienna, Austria, Central Bureau of the Commission Internationale de L'Eclairage, 1986.
- [Chic93] E.J. Chichilnisky, D. Heeger, B.A. Wandell, Functional segregation of color and motion perception examined in motion nulling, *Vision Research*, 33(15):2113, 1993

-
- **[CLSS97]** M.B. Chouikha, G.N. Lu, M. Sedjil, G. Sou, G. Alquie, Buried Triple p-n Junction Structure in a BiCMOS Technology for Color Detection, IEEE BCTM 6.4, pp. 108-111, 1997.
 - **[CM87]** F.W. Campbell, L. Maffei, Kontrast und Raumfrequenz. In Manfred Ritter, Wahrnehmung und visuelles System, pp.132–139, Spektrum der Wissenschaft, 1987.
 - **[Cok87]** D.R. Cok, Signal processing method and apparatus for producing interpolated chrominance values in a sampled color image signal, United States Patent 4,642,678, 1987.
 - **[CSS03]** B. Csatho, T. Schenk, S. Seo, Spectral Interpretation Based on Multisensor Fusion for Urban Mapping, 2nd IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, pp.8-12, 2003.
 - **[DBM83]** H.J.A. Dartnall, J.K. Bowmaker, J.D. Mollon, Microspectrophotometry of human photoreceptors, J.D. Mollon and L.T. Sharpe (editors) Colour Vision, Physiology and Psychophysics, pp.69-80, Academic Press, London, 1983.
 - **[Dem91]** W. Demtröder: Laserspektroskopie, Berlin, Springer, 1991.
 - **[DHBF00]** Doswald, Haflinger, Blessing, Felber, Niederer, Fichtner: A 30-Frames/s Megapixel Real-Time CMOS Image Processor, IEEE Journal of solid-state circuits, 35(11):1732-1743, November 2000.
 - **[DIN5033]** Normenausschuß Farbe, DIN 5033, Teil 1: Farbmessung, Grundbegriffe der Farbmessung, Technischer Report, Deutsches Institut für Normung e.V., März 1979.
 - **[Drep99]** C. Drepper, Farbe macht den Unterschied, <http://www.uni-bielefeld.de/biologie/Kybernetik/pages/volker/bericht1999/farbe.html>
 - **[Dris78]** W.G. Driscoll, Handbook of optics, New York, McGraw Hill, 1978.
 - **[EKV97]** C. Enz, F. Krummenacher, E.A. Vittoz, An Analytical MOS Transistor Model Valid in All Regions of Operation and Dedicated to Low-Power and Low-Current Applications, Analog Integrated Circuits and Signal Processing, 8(1), Juli 1997.
 - **[Fair98]** M. D. Fairchild, Color Appearance Models, Addison-Wesley, 1998.
 - **[Fisch97]** G. Fischer, Modellbildung der Farbproduktion photographischer Filme optoelektronischen innen Abtastsystemen, Shaker Verlag, 1997.
 - **[FHD00]** E.C. Fox, J. Hynecek, D.R. Dykaar, Wide-Dynamic-Range Pixel with Combined Linear and Logarithmic Response and Increased Signal Swing, IS&T/SPIE 12th International Symposium on Electronic Imaging 2000, San Jose, USA, 2000.
 - **[Fos92]** E.R. Fossum, Active pixel sensors (APS) - Are CCDs dinosaurs?, Proc. SPIE, vol. 1900, pp. 214, 1992.
 - **[Fos97]** E. R. Fossum, CMOS image sensors: Electronic camera-on-chip, IEEE Trans. Electron Devices, vol. 44, pp. 1689-1698, Oktober 1997.
 - **[Free88]** Freeman, Median filter for reconstructing missing color samples, United States Patent 4,724,395, 1988.
 - **[FS98]** E.P. Fischer, N. Silvestrini, Farbsysteme in Kunst und Wissenschaft, Köln, 1998.
 - **[GC01]** M. Gupta, T. Chen, Vector color filter array demosaicing, Proceedings of the SPIE Electronic Imaging 2001 conference, Vol. 4306, San Jose, CA, Januar 2001.

-
- [GGAS04] B.K. Gunturk, J. Glotzbach, Y. Altunbasak, R.W. Schafer, R.M. Mersereau, Demosaicking: Color filter array interpolation in single chip digital cameras, IEEE Signal Processing Magazine (Special Issue on Color Image Processing), September 2004.
 - [GH96] K.R. Gegenfurtner, M.J. Hawken, Interaction of motion and color in the visual pathways, Trends in Neurosciences, 19:394-401, 1996.
 - [Gold97] E.B. Goldstein, Wahrnehmungspsychologie, Spektrum Akad. Verlag, 1997.
 - [GP89] A. Gorea, T.V. Papathomas, Motion processing by chromatic and achromatic visual pathways, J. Opt. Soc. Am. A, 6(4):590–602, 1989.
 - [Grau03] A. Graupner, Robuster Entwurf und statistische Modellierung für Bildsensoren mit hochparalleler analoger Bildverarbeitungseinheit, Dissertation, Dresden, 2003.
 - [SGS01] S. Getzlaff, J. Schreiter, A. Graupner, and R. Schüffny: A System-On-Chip Realization of a CMOS Image Sensor with Programmable Analog Image Preprocessing, Proc. IEEE International Symposium on Circuits and Systems ISCAS'2001, 2001.
 - [Gui04] E. Guizzo, Coming Soon: Trillion-Color TV, IEEE Spectrum, pp 12-13, August 2004.
 - [GYF99] A. El Gamal, D. Yang, B. Fowler, Pixel level processing -- Why, what and how?, Proc. SPIE, vol. 3650, pp. 2-13, Jan. 1999.
 - [Haf98] W. Hafner, Segmentierung von Video-Bildfolgen durch Adaptive Farbklassifikation, Dissertation, Institut für Informatik der Technischen Universität München, 1998.
 - [Ham97] F.J. Hamilton, J.E. Adams, Adaptive color plan interpolation in single sensor color electronic camera, United States Patent 5,629,734, 1997.
 - [Har38] Hartline H.K.: The response of single optic nerve fibers of the vertebrate eye to illumination of the retina, American Journal of Physiology, 121, 400-415, 1938.
 - [Hara04] Y. Hara, Matsushita develops hybrid CCD/CMOS sensor, EE Times, 17. Februar 2004.
 - [Hau94] G. Hauske, Systemtheorie der visuellen Wahrnehmung, B. G. Teubner Stuttgart, 1994.
 - [Hau99] R. Hauschild, Integrierte CMOS-Kamerasysteme für die zweidimensionale Bildsensorik, Dissertation, Universität Duisburg, 1999.
 - [HCG95] G. Halder, P. Callerts, W.J. Gehring, New perspectives on eye evolution, Current opinion in Genetics & Development, 5:602-609, 1995.
 - [HDBT97] J.E.D. Hurwitz, P.B. Denyer, D.J. Baxter, G. Townsend, An 800K-pixel color CMOS sensor for consumer still cameras, SPIE proc. ; 3019, S. 115-124, 1997.
 - [Helm24] H.v. Helmholtz. Physiological optics. Rochester, New York: Optical Society of America, 1924.
 - [Her1878] E. Hering. Zur Lehre vom Lichtsinn. Gerold und Söhne, Wien, 1878.
 - [HH96] R. Hogervorst, J.H. Huijsing, Design of low-voltage, low-power operational amplifier cells, Kluwer Academic Publishers, 1996.
 - [HKS99] Herzog, Knipp, Stiebig, König: Characterization of novel three and six channel color moire free sensors, SPIE Vol. 3648, San Jose, California, Jan. 1999.

-
- **[HMS95]** E. Hering, R. Martin, M. Stohrer, Physik für Ingenieure, VDI Verlag, 5. Auflage, 1995.
 - **[HNN98]** S.A. Hagstrom, J. Neitz, M. Neitz, Variations in cone populations for red-green color vision examined by analysis of mrna, Neuroreport, 9(9):1963–7, 1998.
 - **[How79]** M.J. Howes, D.V. Morgan, Charge-Coupled Devices and Systems, John Wiley & Sons, Chichester, 1979.
 - **[Hunt71]** R.W.G. Hunt, Objectives in Color Reproduction. Proc. Inter.Soc. Color Council, pp.1-37, 1971.
 - **[Hunt95]** R.W.G. Hunt, The reproduction of colour, 5th ed. Kingston-upon-Thames: Fountain Press, 1995.
 - **[Hupp00]** J. Huppertz, 2-D CMOS Bildsensorik mit integrierter Signalverarbeitung, Dissertation, Gerhard-Mercator-Universität - Gesamthochschule Duisburg, 2000.
 - **[HYKV00]** W.K. Henson, N. Yang, S. Kubicek, E.M. Vogel, J.J. Wortman K. De Mayer, A. Naem, Analysis of Leakage Currents and Impact on Off-State Power Consumption for CMOS Technology in 100-nm Regime, IEEE Transactions of Electron Devices vol.47, No.7, pp1393-1400, Juli 2000.
 - **[ISO9241]** ISO 9241, Ergonomic requirements for office work with visual display terminals (VDTs). Parts 3, 6, 7, 8.
 - **[Jaco93]** G.H. Jacobs, The distribution and nature of colour vision among the mammals, Biological Reviews, 68:413-471, 1993.
 - **[Jer77]** A.J. Jerri, The Shannon Sampling Theorem - Its Various Extensions and Applications: A Tutorial Review, Proceedings of the IEEE, 65(11):1565–1596, November 1977.
 - **[John25]** J.B. Johnson, The Schottky Effect In Low Frequency Circuits, Physical Review. vol. 26, pp. 71-85, 1925.
 - **[Kel73]** D.H. Kelly, Spatio-temporal frequency characteristics of color-vision mechanisms, J. Physiol., 228:55–72, 1973.
 - **[Kim99]** R. Kimmel, Demosaicing: Image reconstruction from CCD samples, IEEE Transaction Image Processing, vol. 8 pp. 1221-1228, 1999.
 - **[KLLG01]** S. Kleinfelder, S. Lim, X. Liu, A. El Gamal, A 10,000 Frames/s CMOS Digital Pixel Sensor, IEEE Journal of Solid State Circuits, Vol. 36, No. 12, pp.2049-2059, Dezember 2001.
 - **[Kön1894]** A. König. Über den menschlichen Sehpurpur und seine Bedeutung für das Sehen. Akademie der Wissenschaften, Sitzungsberichte, pp.577-598. 1894.
 - **[Kön97]** F. König, Reconstruction of Natural Spectra from a Color Sensor Using Nonlinear estimation methods, Proc. IS&T's 50th Annual Conference, Cambridge, MA, USA, pp. 454-458, 1997.
 - **[Krue93/1]** H. Krueger, Arbeit mit dem Bildschirm, Handbuch der Arbeitsmedizin, Konietzko, Dupuis (Hrsg.) Kap. IV-9.2.1, 1993.
 - **[Krue93/2]** H. Krueger, Arbeiten mit dem Bildschirm - aber richtig! Bayerisches Staatsministerium für Arbeit, Familie und Sozialordnung, 1993.

-
- **[Lang95]** H. Lang, Farbwiedergabe in den Medien. Muster Schmidt Verlag, Göttingen, Zürich, 1995.
 - **[Lee96]** B.B. Lee, Receptive field structure in the primate retina, Vision Research, 36(5), pp.631–44, 1996.
 - **[LP94]** A.C. Laroche, M.A. Prescott, Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients, United States Patent 5,373,322, 1994.
 - **[Luth27]** R. Luther, Aus dem Gebiet der Farbreizmetrik, Zeitschrift für technische Physik, Nr.12, pp.540 - 558, 1927.
 - **[MacA42]** D.L. MacAdam, Visual sensitivities to color differences in daylight. Journal Optical Society of America, 32, pp. 247-274, 1942.
 - **[Mar94]** J. Marchesi, Professionelle Beleuchtungstechnik, Verlag Photographie, 1994.
 - **[Max1856]** J.C. Maxwell, Theory of the perception of colors, Transactions of the Royal Scottish Society of Arts, 4, pp.394-400. Nachdruck aus D.L. MacAdam (Ed.), Sources of color science pp. 62-66, Cambridge, MA, MIT Press, 1856/1970.
 - **[Max1857]** J.C. Maxwell, The diagram of colors, Transactions of the Royal Society of Edinburgh, 21, pp.275-298. Nachdruck aus D.L. MacAdam (Ed.), Sources of color science pp. 67-72. Cambridge, MA, MIT Press. 1857/1970.
 - **[Maz04]** Mazet, True Colour Farbsensoren nach DIN 5033, http://www.mazet.de/mazet_d1/news_d/presse/true_colour_sensor.html, 2004.
 - **[MBM81]** K.D. Mörike, E. Betz, W. Mergenthaler, Biologie des Menschen. Quelle&Meyer, 1981.
 - **[Mic04]** Micron Technology, Inc, MT9V011 1/4-INCH VGA CMOS ACTIVE-PIXEL DIGITAL IMAGE SENSOR, Micron Technology, Inc. 2004.
 - **[Mö92]** A. Möschwitzer: Grundlagen der Halbleiter- & Mikroelektronik. Band 1: Elektronische Halbleiterbauelemente, Carl Hanser Verlag, 1992.
 - **[MP02]** D.D. Muresan, T.W. Parks, Optimal Recovery Demosaicing, IASTED Signal and Image Processing, Hawaii, 2002.
 - **[MWW]** <http://www.m-ww.de/krankheiten/augenkrankheiten/farbsehstoerungen.html>
 - **[Nie82]** K. Niederdrenk, Die endliche Fourier- und Walsh-Transformation mit einer Einführung in die Bildverarbeitung, Friedr. Vieweg & Sohn, 1982.
 - **[NN02]** A.M.I. Nilsson, P.U.W. Nordblom, Interpolation in Color Filter Arrays, Master Thesis, Malmö University, School of Technology and Society, 2002.
 - **[Nob68]** P. Noble, Self-scanned image detector arrays, IEEE Transactions Electron Devices, vol. ED-15, p.202, 1968.
 - **[NPNF97]** J. Nakamura, B. Pain, T. Nomoto, T. Nakamura, E.R. Fossum: On-Focal-Plane Signal Processing for Current-Mode Active Pixel Sensors, IEEE Transactions on Electron Devices, 44(10), pp1747-1758. Oct. 1997.
 - **[OV96]** D. Osorio, M. Vorobyev, Colour vision as an adaptation to frugivory in primates, Proceedings of the Royal Society (London) B, 263:593-599, 1996.

-
- **[PS86]** J. Pokorny, V.C. Smith, Colorimetry and color discrimination. K.R. Boff, L. Kaufmann, J.P. Thomas (Editors), Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception. John Wiley & Sons, New York, 1986.
 - **[Poy95]** C. Poynton, A Guided Tour of Color Space, New Foundations for Video Technology, Proceedings of the SMPTE Advanced Television and Electronic Imaging Conference, San Francisco, pp.167-180, Februar 1995.
 - **[Poy96]** C. Poynton, A Technical Introduction to Digital Video, John Wiley & Sons, New York, 1996.
 - **[Poy97]** C. Poynton, Poynton's Color FAQ, <http://www.gimlay.org/~andoh/cg/faq/Color-FAQ.html>, 1997.
 - **[Püg01]** M. Pügner, Entwicklung und Hardwareimplementierung eines Farbinterpolationsalgorithmus für Bildsensoren mit Farbmosaikfiltern, Diplomarbeit, TU Dresden, 2001.
 - **[RB64]** W.A.H. Rushton, H.D. Baker, Red/green sensitivity in normal vision, Vision Research, 4(1):75–85, 1964.
 - **[RD92]** N. Ricquier, B. Dierickx, Pixel structure with logarithmic response for intelligent and flexible imager architectures, Microelectron. Engineering, vol. 19, pp. 631:634, 1992.
 - **[RGH98]** Ramacher, Geib, Heer, Kodytek, Koren, Werner, Dohndorf, Schlüßler, Poidevin, Kirmser, 1-Chip-Implementation einer digitalen CMOS-Video-Kamera mit 110dB Dynamik, Mikroelektronik für die Informationstechnik, ITG, pp.311-317, Hannover, 1998.
 - **[Rich81]** M. Richter, Einführung in die Farbmatrik, de Gruyter, Berlin, New York, 1981.
 - **[RK GK99]** U. Ramacher, I. Koren, H. Geib, S. Kirmser, C. Heer, J.-U. Schlüßler, J. Dohndorf, J. Werner, Single-Chip Video Camera with Multiple Integrated Functions. In International Solid-State Circuits Conference ISSCC'99, pp.306-307, 1999.
 - **[Rock85]** I. Rock, Wahrnehmung, Spektrum Akademischer Verlag, 1985.
 - **[Rush72]** W.A.H. Rushton, Visual pigments in man, H. Autrum (Series Ed.) & H.J.A. Dartnall (Ed.), Handbook of sensory physiology, Vol. VII/1, Photochemistry of vision, pp.364-394, Springer, New York, 1972.
 - **[Rull02]** M. Rullmann: Entwurf und Implementierung von Algorithmen für die Korrektur von Bildfehlern bei CMOS-Sensoren, Diplomarbeit, Dresden, 2002.
 - **[SCB88]** M.C. Stone, W.B. Cowan, J.C. Beatty, Color Gamut Mapping and the Printing of Digital Color Images, in ACM Transactions on Graphics, Vol. 7, No. 3, Oktober 1988.
 - **[Schäp93]** K. Schäpfer, Farbmatrik in der Reproduktionstechnik und im Mehrfarbendruck, St. Gallen, 1993.
 - **[Schie01]** Ch. Schierz, Sehen und Bildschirm, Eidgenössische Technische Hochschule Institut für Hygiene und Arbeitsphysiologie, 2001.
 - **[Schick01]** P. Schicketanz, Untersuchung von Technologien für Farbfilter und Mikrolinsen zur monolithischen Integration in CMOS-Fotosensorarrays, Dissertation, Gesamthochschule Duisburg, 2001.
 - **[Schm95]** R.F. Schmidt, Neuro- und Sinnesphysiologie, 2. Aufl. Springer, Berlin, 1995.
 - **[Schot26]** W. Schottky, Small-Shot Effect And Flicker Effect, Physical Review. vol. 28, pp.74-103, 1926.

-
- **[SG03]** K. Salama, A. El Gamal, Analysis of Active Pixel Sensor Readout Circuit, IEEE Transactions Circuits and Systems I: Fundamental Theory and Applications, Vol. 50, No. 7, p.941-945, Juli 2003.
 - **[SGGS00]** J. Schreiter, A. Graupner, S. Getzlaff, R. Srowik, R. Schüffny, A CMOS Image Sensor with Parallel Analog Processing Unit for Transformation and Spatial Convolutions, 4th International Multiconference on Systemics, Cybernetics and Informatics SCI'2000, volume 5, pp.298-301, Orlando, Florida, Juli 2000.
 - **[SGSG01]** J. Schreiter, A. Graupner, R. Srowik, S. Getzlaff, R. Schüffny: A Multi-Input Pipelined A/D Converter for Special Signal Processing Applications, Mixed Design of Integrated Circuits and Systems (MIXDES), Zakopane, Juni 2001.
 - **[SK95]** C.F. Stromeyer, R.E. Kronauer, A. Ryu, A. Chaparro, R.T. Eskew Jr., Contributions of human long-wave and middle-wave cones to motion detection. J. Physiol., 485:221–43, 1995.
 - **[SKB87]** J.L. Schnapf, T.W. Kraft, D.A. Baylor, Spectral sensitivity of human cone photoreceptors, Nature, 325:439–441, 1987.
 - **[Sony03]** Sony Corporation, Realization of natural color reproduction in Digital Still Cameras, closer to the natural sight perception of the human eye, Press Release, <http://www.sony.net/SonyInfo/News/Press/200307/03-029E/>, 16. Juli 2003.
 - **[Stein93]** R. Steinbrecher, Bildverarbeitung in der Praxis, München, Wien, Oldenbourg, 1993.
 - **[Sva56]** G. Svaetichin, Spectral responses from single cones, Acta Physiol Scand 39 (Suppl 134), pp.17-46, 1956.
 - **[TB88]** D.-Y. Tzeng, R.S. Berns, Spectral-Based Ink Selection for Multiple-Ink. Printing I. Colorant Estimation of Original Object, Proc. 6th IS&T/SID Color Imaging Conference, in progress, 1988.
 - **[TBK96]** D.P. Triantis, A.N. Birbas, D. Kondis, Thermal noise modeling for short-channel MOSFET's, IEEE Trans. Electron Devices, Vol. 43, no. 11, pp.1950-1955, November 1996.
 - **[TFG01]** H. Tian, B. Fowler, A. El Gamal, Analysis of Temporal Noise in CMOS APS, IEEE Journal of Solid State Circuits, Vol. 36, No.1, pp. 92-101, Januar 2001.
 - **[TG01]** Hui Tian, Abbas El Gamal, Analysis of 1/f Noise in Switched MOSFET Circuits, IEEE Transactions On Circuits and Systems--II, Vol. 48, No. 2, Februar 2001.
 - **[Theu95]** A.J.P. Theuwissen, Solid-State Imaging with Charge-Coupled Devices, Kluwer Academic Publishers, Dordrecht, 1995.
 - **[Tian00]** Hui Tian, Noise Analysis in CMOS Image Sensors, Dissertation, Stanford University, 2000.
 - **[TS90]** L.G. Thorell, W.J. Smith, Using computer color effectively, Prentice-Hall, 1990.
 - **[Vitt96]** E.A. Vittoz, Analog VLSI implementation of neural networks, Handbook of Neural Computation, Ltd and Oxford University Press, 1996.
 - **[Wen03]** P. Wenig, Menschliche Wahrnehmung bei der Quellencodierung, Seminar Digitale Signalverarbeitung in Multimedia-Geräten, 2003.
 - **[WGG02]** B. Wandell, A.El Gamal, B. Girod, Common Principles of Image Acquisition Systems and Biological Vision, Proceeding of IEEE, Vol. 90, No. 1, pp.5-17, Januar 2002.

-
- [WMES99] M. Wang, T.L. Mayhugh, Jr., S.H.K. Embabi, E. Sanchez-Sinencio, Constant-gm Rail-to-Rail CMOS Op-Amp Input Stage with Overlapped Transition Regions, IEEE Journal of Solid-State Circuits, vol.34, no.2, 1999.
 - [WMH81] D.R. Williams, D.I.A. MacLeod, M. Hayhoe, Punctate sensitivity of the blue-sensitive mechanisms, Vision Research, 21(9):1357–75, 1981.
 - [WS04] S. Winkler, S. Süssstrunk, Visibility of noise in natural images, Proc. IS&T/SPIE Electronic Imaging 2004: Human Vision and Electronic Imaging IX, IS&T/SPIE Vol. 5292, pp.121-129, 2004.
 - [WS82] Günter Wyszecki, W.S. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formula, John Wiley & Sons, New York, 2nd edition, 1982.
 - [YES97] F. You, S.H.K. Embabi, E. Sanchez-Sinencio, On the Common Mode Rejection Ratio in Low Voltage Operational Amplifiers with Complementary N-P Input Pair, IEEE Trans. on Circuits and Systems II, Vol. 44, no. 8, pp. 678-683, August 1997.
 - [YFG99] D. Yang, B. Fowler, A. El Gamal, A Nyquist Rate Pixel Level ADC for CMOS Image Sensors, In IEEE Journal of Solid State Circuits, Vol.34, No.3, pp.348-356, March 1999.
 - [YR99] S. Yokoyama, F.B. Radlwimmer, The Molecular Genetics of Red and Green Color Vision in Mammals, Genetics Society of America, 1999.
 - [YTO02] M. Yamaguchi, T. Teraji, K. Ohsawa, T. Uchiyama, H. Motomura, Y. Murakami, N. Ohya, Color image reproduction based on the multispectral and multiprimary imaging: Experimental evaluation, Color Imaging: Device Independent Color, Color Hardcopy and Applications VII, Proc. SPIE, Vol.4663, pp.15-26, 2002.
 - [Young1802] T. Young, On the theory of light and colours, Philosophical Transactions, pp.12–48, 1802.

Eigene Publikationen als Hauptautor

- [Hen04/1] S. Henker, Hierarchical Simulation of Pulse-coupled Neural Networks. In 2004 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS 2004), Hiroshima, Japan, Jul, 2004.
- [Hen04/2] S. Henker, Design von Multikanal-CMOS-Sensoren. Institutskolloquium „Integrierte Schaltungen in speziellen Anwendungen der Signalverarbeitung“, Rostock, Mai 2004.
- [HSS04] S. Henker, J.-U. Schlüßler, R. Schüffny, Design von Multikanal-CMOS-Sensoren, In Dresdner Arbeitstagung Schaltungs- und Systementwurf DASS'2004, 2004.
- [HSS03] S. Henker, J.-U. Schlüßler, R. Schüffny, Concept of Color Correction on Multi-Channel CMOS Sensors, In Digital Image Computing: Techniques and Applications, Proc. of the 7th Biennial Australian Pattern Recognition Society Conference - DICTA 2003, volume 2, pp.771-780, Macquarie University, Sydney, Australia, Dezember 2003.
- [Hen03] S. Henker, Verbesserung der Farbwertwiedergabe mit Multi-Kanal-CMOS-Photosensoren. In ITG-Fachgruppensitzung "Mikroelektronik neuronaler Netze", Dresden, September 2003.
- [HSS01] S. Henker, J.-U. Schlüßler, R. Schüffny, Algorithmic Infrared Correction for CMOS Color Sensors. In 8th Microoptics Conference, Osaka, Japan, 2001.
- [HGG01] S. Henker, S. Getzlaff, A. Graupner, J. Schreiter, R. Schüffny. Systemsimulation von hochparallelen Rechenfeldern am Beispiel eines Sensors mit integrierter Vorverarbeitung. In 10. E.I.S.-Workshop Entwurf Integrierter Schaltungen und Systeme, pp.123-126, Dresden, April 2001.
- [HGG00/1] S. Henker, S. Getzlaff, A. Graupner, J. Schreiter, R. Schüffny. Einsatz eines hierarchischen Simulators zur Analyse von hochparallelen analogen Rechenfeldern. In ITG-Fachgruppensitzung, Dresden, Juli 2000.
- [HGG00/2] S. Henker, A. Graupner, S. Getzlaff, J. Schreiter, R. Schüffny. A New Hierarchical Simulator for Highly Parallel Analog Processor Arrays. In R.Merker and W.Schwarz (Editors), System Design Automation -Fundamentals, Principles, Methods, Examples, pp.176-182. Kluwer Academic Publishers, 2000.

Publikationen als Co-Autor

- [GFH03] S. Getzlaff, M. Franz, S. Henker, A. Graupner, G. Fettweis, R. Schüffny, Mixed-Signal-Implementierung eines Rake-Empfängers für UMTS-Systeme. In 11. Workshop Entwurf integrierter Schaltungen und Systeme EIS'2003, pp.111-116, März 2003.
- [GGS01] A. Graupner, S. Getzlaff, J. Schreiter, S. Henker, R. Schüffny, Systematischer Systementwurf eines CMOS-Bild-Sensors mit hochparalleler analoger Vorverarbeitung. In 10. E.I.S.-Workshop Entwurf Integrierter Schaltungen und Systeme, pp.11-16, Dresden, April 2001.

Patent

- S. Henker, R. Schüffny, Schaltungsanordnung zum Erstellen eines Anfangswertes an einem ladungsspeichernden Element, Aktenzeichen AZ 100620-2631-04220, 2004.

Anhang A

Physikalischer und mathematischer Anhang

1 Erläuterung physikalischer Größen

1.1 Sichtbares Licht

Sichtbares Licht ist ein kleiner Ausschnitt im elektromagnetischen Spektrum (380 nm = violettes Licht - 750 nm = rotes Licht), der von ultravioletter (UV) und infraroter Licht (IR) begrenzt wird. Das in der Wahrnehmung weiße Licht ist dabei eine Zusammensetzung von Licht verschiedener Wellenlängen zwischen rot und violett.

Die Beschreibung der Eigenschaften von Licht kann auf verschiedene Arten geschehen, die Teilaspekte jeweils besonders hervorheben. In der geometrischen Optik sind dies geradlinige Ausbreitung, Reflexion und Brechung. Daraus abgeleitet sind zum Beispiel die Abbildungseigenschaften optischer Geräte beschreibbar. Das Anschauungsmodell der Lichtstrahlen ist dabei nur anwendbar, wenn die Dimensionen nicht in der Größenordnung der Wellenlänge liegen. An dieser Stelle setzt die Wellenoptik an, mit der sich Interferenz- und Beugungserscheinungen sowie das Auflösungsvermögen optischer Geräte, Holographie etc. erklären lassen.

Das Wellenmodell beschreibt das Licht dabei als eine elektromagnetische Welle, die sich von jedem Punkt in alle Richtungen ausbreitet. Die Richtungen des Wellenvektors \mathbf{k} , der elektrischen Feldstärke \mathbf{E} und der magnetischen Induktion \mathbf{B} sind orthogonal. Der Vektor der elektrischen Feldstärke ist dabei in einer Ebene konstant. Licht ist somit linear polarisiert. Die Ebene von \mathbf{k} - und \mathbf{E} -Vektor ist die Schwingungsebene und die \mathbf{k} - \mathbf{B} -Ebene ist die Polarisationssebene. Monochromatisches Licht wird durch die Frequenz f bzw. die Wellenlänge λ beschrieben. Die Größen sind durch die Beziehung $f \cdot \lambda = c$ verknüpft, dabei ist c die Ausbreitungsgeschwindigkeit des Lichts im Vakuum.

In der Quantenoptik schließlich werden Phänomene, wie Photoeffekt und Comptoneffekt beschreibbar. Das Licht der Frequenz f besteht in diesem Modell aus einzelnen Lichtquanten, sogenannten Photonen, der Energie $h \cdot f$, die sich mit der Lichtgeschwindigkeit c geradlinig ausbreiten. h ist dabei das Plancksche Wirkungsquantum. Die einheitliche Charakterisierung von Licht ist dabei in allen drei Beschreibungsmodellen durch die Frequenz beziehungsweise die Wellenlänge möglich. Im Strahlenmodell kann die Wellenlänge als Farbe und im Quantenmodell als Energie der Lichtquanten interpretiert werden.

1.2 Radiometrische und Photometrische Grundgrößen

1.2.1 Bogenmaß b , Raumwinkel ω

Das Bogenmaß $m = b/r$ ist das Verhältnis der überdeckten Strecke b zum Radius r . Die Einheit des Bogenmaßes ist Radiant [rad]. Es gilt stets $0 \leq b < 2\pi$ [rad].

Der Raumwinkel $\omega = A/r^2$ ist eine räumliche Erweiterung des zweidimensionalen Winkels im Bogenmaß. Er ist definiert durch das Verhältnis der überdeckten Kugeloberfläche A zum Quadrat des Kugelradius r . Die Einheit des Raumwinkels ist Steradian [sr]. Der volle Raumwinkel entspricht per Definition der Fläche einer Einheitskugel 4π [sr].

1.2.2 Strahlung, Lichtmenge

Die Strahlungsenergie Q , gemessen in Joule [J], ist die grundlegende Größe der Radiometrie. Sie beschreibt die Energieabstrahlung einer Strahlungsquelle. Ist die Kennzeichnung als spektrale, auf die Wellenlänge bezogene Größe notwendig, wird oft ein hochgestellter Index λ und

die Angabe von Q^λ verwendet.

Das photometrische Gegenstück zur Strahlungsenergie ist die Lichtmenge $Q(\lambda)$. Die Lichtmenge ist die mit dem spektralen Hellempfindlichkeitsgrad der menschlichen Wahrnehmung bewertete spektrale Strahlungsmenge Q^λ . Die Einheit der Lichtmenge ist Lumensekunde [lm·s] oder Lumenstunde [lm·h].

1.2.3 Strahlungsleistung, Lichtstrom

Eine Strahlungsquelle gibt die Strahlungsleistung $\Phi = dQ/dt$, gemessen in Watt [W], in den Raum ab. Mit hochgestelltem Index λ und der Angabe von Φ^λ wird die spektrale Strahlungsleistung bezeichnet. Die Funktion der spektralen Strahlungsleistung über der Wellenlänge, das Leistungsdichtespektrum der Strahlung, wird hier mit $\phi(\lambda)$ bezeichnet.

In Bezug auf die menschliche Wahrnehmung gibt eine Lichtquelle den Lichtstrom $\Phi(\lambda) = d(Q(\lambda))/dt$ ab; $\Phi(\lambda)$ ist dabei ein Maß für die Leistung der Lichtquelle. Die Einheit für den Lichtstrom ist das Lumen [lm].

1.2.4 Strahlstärke, Lichtstärke

Die Strahlungsleistung pro Einheitsraumwinkel $I = d\Phi/d\omega$, gemessen in Watt pro Steradian [$W \cdot sr^{-1}$] heißt Strahlstärke.

Die Lichtstärke $I(\lambda)$ stellt das photometrische Gegenstück zur Strahlstärke I dar. Sie verbindet die physikalische Größe mit der menschlichen Hellempfindlichkeit. Die Lichtstärke $I(\lambda) = d\Phi(\lambda)/d\omega$ ist der Quotient aus dem von einer Lichtquelle in einer bestimmten Richtung ausgesandten Lichtstrom $\Phi(\lambda)$ und dem durchstrahlten Raumwinkel ω . Seit 1967 ist die Einheit der standardisierten Größe das Candela [cd]. Ein Candela ist dabei die Lichtstärke in einer bestimmten Richtung einer Lichtquelle, die Strahlung einer Wellenlänge von 555 nm aussendet und deren Strahlstärke in dieser Richtung $1/683 W \cdot sr^{-1}$ beträgt. Seit Oktober 1979 orientiert sich die Definition an dem Licht von erstarrtem Platin.

Die Hellempfindlichkeit des Menschen ist nicht für alle Wellenlängen gleich. Ihr relativer Verlauf ist als spektraler Hellempfindlichkeitsgrad $V(\lambda)$ für Tagsehen und $V'(\lambda)$ für Nachtsehen in DIN 5031 definiert. Mit dem photometrischen Strahlungsäquivalent $K(\lambda)$ ist die Berechnung von Absolutwerten möglich, wobei $K(\lambda) = V(\lambda) \cdot K(\lambda)_{\max}$ gilt; mit den Maxima der Hellempfindlichkeit für Hellsicht $K(\lambda)_{\max} = 683 \text{ lm} \cdot W^{-1}$ und $K'(\lambda)_{\max} = 1699 \text{ lm} \cdot W^{-1}$ für Dunkelsicht. Das Maximum des spektralen Hellempfindlichkeitsgrades für Tagsehen liegt bei einer Wellenlänge von 555 nm, für Dunkelsicht bei 508 nm. Das hellsichtige Auge ist beispielsweise für Karminrot (750 nm) ca. 10^4 mal weniger empfindlich als für Grün (555 nm). Eine grüne Lampe mit 100 W wird somit genauso hell wahrgenommen wie eine karminrote Lampe mit 10^6 W. Die Lichtstärke einer Lichtquelle ist weiterhin im allgemeinen winkelabhängig. Die Abstrahlcharakteristik einer Leuchte wird durch die Lichtstärkeverteilungskurve (LVK) angegeben. Für die spektrale Strahlstärke I^λ müßte die Verteilungskurve theoretisch für jede Wellenlänge angegeben werden. In der Praxis werden für Lichtquellen jedoch oft nur das Spektrum und eine LVK angegeben. Isotrope Materialien zeigen dabei ein rotationssymmetrisches Abstrahlverhalten. Die LVK ist dann nur von einem Winkel abhängig. Für anisotrope Materialien muß die LVK in Abhängigkeit von zwei Austrittswinkeln beschrieben werden.

1.2.5 Bestrahlungsstärke, Beleuchtungsstärke

Das Maß für die pro Fläche auftreffende Strahlungsleistung Φ ist die Bestrahlungsstärke $E = d\Phi/dA$, gemessen in Watt pro Quadratmeter [$\text{W}\cdot\text{m}^{-2}$].

Das photometrische Äquivalent ist die Beleuchtungsstärke $E(\lambda) = d\Phi(\lambda)/dA$. Das Maß der Beleuchtungsstärke ist Lux [lx]. Ein Lux entspricht dabei einem Lumen pro Quadratmeter.

1.2.6 Photonenflußdichte

Als Photonenflußdichte Π , gemessen in Anzahl von Photonen pro Quadratmeter und Sekunde [$\text{m}^{-2}\cdot\text{s}^{-1}$], bzw. spektrale Photonenflußdichte Π^λ wird die Quantelung von E bzw. E^λ in einzelne Photonen bezeichnet. E^λ und Π^λ sind über die Beziehung verbunden:

$$\Pi^\lambda = \frac{1}{h \cdot f} \cdot E^\lambda = \frac{\lambda}{h \cdot c} \cdot E^\lambda \quad (\text{A.1})$$

mit dem Planckschen Wirkungsquantum h , der Lichtgeschwindigkeit c und der Frequenz der Strahlung f bzw. deren Wellenlänge λ . Die Funktion der Photonenflußdichte über der Wellenlänge wird hier in Analogie zum Leistungsdichtespektrums mit $\pi(\lambda)$ bezeichnet.

1.2.7 Belichtung

Die Belichtung $H = \int E(\lambda) \cdot dt$ ist ein Maß für die Intensität eines Lichtreizes. Die Einheit der Belichtung ist die Luxsekunde [$\text{lx}\cdot\text{s}$].

1.2.8 Strahldichte und Leuchtdichte

Strahldichte (eng. *radiance*) ist eine wichtige Größe in Bezug auf die Wahrnehmung, da ihr photometrisches Pendant vom menschlichen Auge als Helligkeit wahrgenommen wird. Dabei kommt es nicht auf die empfangene Gesamtmenge an Licht an, um verschiedene Gegenstände als gleich hell zu empfinden, sondern auf deren Strahldichte, dem Verhältnis der Strahlstärke der Lampe zur wirksamen Senderfläche. So ist die Strahldichte L als Strahlungsleistung Φ eines Flächenelementes dA in Richtung eines infinitesimalen Raumwinkels $d\omega$ definiert. Steht das Flächenelement dA dabei nicht senkrecht zu der betrachteten Strahlungsrichtung, ist nur die Flächenprojektion $dA \cdot \cos(\alpha)$ wirksam. Allgemein gilt $L = d^2\Phi / (dA \cdot d\omega \cdot \cos(\alpha))$. Die Einheit der Strahldichte ist [$\text{W}\cdot\text{sr}^{-1}\cdot\text{m}^{-2}$]. Die Strahldichte selbst ist invariant und bleibt in Richtung eines Lichtstrahls konstant. Dies gilt auch im Zusammenhang mit optischen Abbildungen unter der Voraussetzung, daß keine Verluste durch Absorption oder Streuung auftreten. Die Strahldichte, die ein Flächenelement dA_1 in Richtung eines Flächenelements dA_2 verläßt, ist gleich der Strahldichte, die bei dA_2 aus Richtung von dA_1 auftrifft.

Im photometrischen Äquivalent, der Leuchtdichte, ist analog dazu beschrieben. Wird eine Lichtquelle aus einer bestimmten Richtung betrachtet, so erscheint sie desto heller, je kleiner die leuchtende Fläche A bei konstanter Lichtstärke $I(\lambda)$ ist. Die Definition der Leuchtdichte ist daher $L(\lambda) = d^2\Phi(\lambda) / (dA \cdot d\omega \cdot \cos(\alpha))$. Die Einheit für die Leuchtdichte ist Candela pro Quadratmeter [$\text{cd}\cdot\text{m}^{-2}$]. Die Leuchtdichte $L(\lambda)$ ist ein Maß für die subjektiv gesehene Helligkeit. Ein weißes Papier auf einem gut beleuchteten Büroschreibtisch hat etwa eine Leuchtdichte von $100 \text{ cd}\cdot\text{m}^{-2}$; die Leuchtdichte einer Leuchtstofflampe beträgt ca. $14'000 \text{ cd}\cdot\text{m}^{-2}$.

Ein spezieller Strahler ist der Lambert-Strahler, der sich dadurch auszeichnet, daß seine Strahldichte unter jedem Betrachtungswinkel konstant ist und deshalb von jeder Blickrichtung gleich

hell erscheint. Die sichtbare Fläche nimmt dabei im gleichen Verhältnis ab wie die Lichtstärke.

2 Mathematischer Anhang

2.1 Funktionsapproximation mit der Methode des kleinsten quadratischen Fehlers

Eine Funktion $a_i(\lambda)$ soll durch Linearkombination von Basisfunktionen $s_1(\lambda) \dots s_n(\lambda)$ in der Form $a_i'(\lambda) = c_{i1} \cdot s_1(\lambda) + \dots + c_{in} \cdot s_n(\lambda)$ nachgebildet werden. Dabei sei der quadratische Fehler der Approximation, bestimmt durch $F = \int_a^b (a_i(\lambda) - a_i'(\lambda))^2 \cdot d\lambda$ im Intervall $[a, b]$ zu minimieren. Der Quadratfehler und dessen partiellen Ableitungen nach den Koeffizienten c_{ij} ergeben sich zu:

$$F = \int_a^b (a_i(\lambda) - (c_{i1} \cdot s_1(\lambda) + \dots + c_{in} \cdot s_n(\lambda)))^2 \cdot d\lambda \quad (\text{A.2})$$

$$\frac{\delta}{\delta c_{ij}} F = \int_a^b -2 \cdot (a_i(\lambda) - (c_{i1} \cdot s_1(\lambda) + \dots + c_{in} \cdot s_n(\lambda))) \cdot s_j(\lambda) \cdot d\lambda \quad (\text{A.3})$$

$$\frac{\delta^2}{\delta c_{ij}^2} F = \int_a^b 2 \cdot s_j(\lambda) \cdot s_j(\lambda) \cdot d\lambda \quad (\text{A.4})$$

Die zweite Ableitung Gleichung (A.4) ist stets positiv, somit sind die Nullstellen der ersten Ableitung nach Gleichung (A.3) Minima des Quadratfehlers entsprechend der Aufgabenstellung.

$$\frac{\delta}{\delta c_{ij}} F = \int_a^b -2 \cdot (a_i(\lambda) - (c_{i1} \cdot s_1(\lambda) + \dots + c_{in} \cdot s_n(\lambda))) \cdot s_j(\lambda) \cdot d\lambda = 0 \quad (\text{A.5})$$

$$\int_a^b a_i(\lambda) \cdot s_j(\lambda) \cdot d\lambda = \int_a^b (c_{i1} \cdot s_1(\lambda) + \dots + c_{in} \cdot s_n(\lambda)) \cdot s_j(\lambda) \cdot d\lambda \quad (\text{A.6})$$

Für $j = 1 \dots n$ wird aus Gleichung (A.6) ein Gleichungssystem mit n Gleichungen in der Form:

$$\begin{aligned} \int_a^b a_i(\lambda) \cdot s_1(\lambda) \cdot d\lambda &= c_{i1} \cdot \int_a^b s_1(\lambda) \cdot s_1(\lambda) \cdot d\lambda + \dots + c_{in} \cdot \int_a^b s_n(\lambda) \cdot s_1(\lambda) \cdot d\lambda \\ &\vdots \\ \int_a^b a_i(\lambda) \cdot s_n(\lambda) \cdot d\lambda &= c_{i1} \cdot \int_a^b s_1(\lambda) \cdot s_n(\lambda) \cdot d\lambda + \dots + c_{in} \cdot \int_a^b s_n(\lambda) \cdot s_n(\lambda) \cdot d\lambda \end{aligned} \quad (\text{A.7})$$

Die Überführung in Vektorschreibweise mit dem Vektor der Basisfunktionen $\underline{s}(\lambda) = [s_1(\lambda) \dots s_n(\lambda)]^T$ und dem Koeffizientenvektor $\underline{c}_i = [c_{i1} \dots c_{in}]$ führt zu der verkürzten Darstellung mit:

$$\int_a^b a_i(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda = \int_a^b \underline{s}(\lambda) \cdot \underline{s}(\lambda)^T \cdot \underline{c}_i^T \cdot d\lambda = \underline{c}_i \cdot \int_a^b \underline{s}(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda \quad (\text{A.8})$$

Wird nicht nur eine Funktion $a_i(\lambda)$ nachgebildet, sondern mehrere $a_i(\lambda)$ mit $i = 1 \dots m$ und $\underline{a}(\lambda) = [a_1(\lambda) \dots a_m(\lambda)]^T$, so ergibt sich:

$$\begin{aligned} \int_a^b a_1(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda &= \underline{c}_1 \cdot \int_a^b \underline{s}(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda \\ &\vdots \\ \int_a^b a_m(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda &= \underline{c}_m \cdot \int_a^b \underline{s}(\lambda) \cdot \underline{s}(\lambda)^T \cdot d\lambda \end{aligned} \quad (\text{A.9})$$

und der Vektorschreibweise:

$$\int_a^b \underline{a}(\lambda) \cdot \underline{s}(\lambda)^\top \cdot d\lambda = \mathbf{G} \cdot \int_a^b \underline{s}(\lambda) \cdot \underline{s}(\lambda)^\top \cdot d\lambda \quad (\text{A.10})$$

$$\text{mit: } \mathbf{G} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix}$$

Zu einer äquivalenten Lösung gelangt man durch direkte Ableitung nach der Matrix \mathbf{G} in der Vektordarstellung des quadratischen Fehlers. Der Quadratfehler in Vektorschreibweise sei:

$$E = \int_a^b (\underline{a}(\lambda) - \underline{a}'(\lambda))^2 \cdot d\lambda = \int_a^b (\underline{a}(\lambda) - \mathbf{G} \cdot \underline{s}(\lambda))^2 \cdot d\lambda \quad (\text{A.11})$$

Daraus ergeben sich:

$$[E] = \int_a^b (\underline{a}(\lambda) - \mathbf{G} \cdot \underline{s}(\lambda)) \cdot (\underline{a}(\lambda) - \mathbf{G} \cdot \underline{s}(\lambda))^\top \cdot d\lambda \quad (\text{A.12})$$

und die erste und zweite Ableitung nach der Koeffizientenmatrix \mathbf{G} zu:

$$\left[\frac{\delta}{\delta \mathbf{G}} \cdot E \right] = \int_a^b 2 \cdot (\underline{a}(\lambda) - \mathbf{G} \cdot \underline{s}(\lambda)) \cdot (-\underline{s}^\top(\lambda)) \cdot d\lambda \quad (\text{A.13})$$

$$\left[\frac{\delta^2}{\delta \mathbf{G}^2} \cdot E \right] = \int_a^b 2 \cdot \underline{s}(\lambda) \cdot \underline{s}(\lambda)^\top \cdot d\lambda \quad (\text{A.14})$$

Die Extrema des Quadratfehlers ergeben sich an den Nullstellen der ersten Ableitung zu:

$$\left[\frac{\delta}{\delta \mathbf{G}} \cdot E \right] = \int_a^b 2 \cdot (\underline{a}(\lambda) - \mathbf{G} \cdot \underline{s}(\lambda)) \cdot (-\underline{s}^\top(\lambda)) \cdot d\lambda = [0] \quad (\text{A.15})$$

und führt umgestellt zu der allgemeinen Lösung:

$$\int_a^b \underline{a}(\lambda) \cdot \underline{s}(\lambda)^\top \cdot d\lambda = \mathbf{G} \cdot \int_a^b \underline{s}(\lambda) \cdot \underline{s}(\lambda)^\top \cdot d\lambda \quad (\text{A.16})$$

Aus der positiven zweiten Ableitung folgt, daß die Extrema nach Gleichung (A.16) Minima gemäß der Aufgabenstellung sind.

3 Modellierungen

3.1 MOS-Transistormodell

Als Transistormodell für MOS-Transistoren wird das Modell nach Enz et al. [EKV97] verwendet. Dieses besitzt getrennte Funktionen für *weak-inversion* und *strong-inversion*. Der Übergang zwischen beiden Arbeitsbereichen wird durch eine Interpolation realisiert. Der Vorteil des Modells liegt im besonderen in der Symmetrie von Drain und Source. Eine Besonderheit liegt in der Notation der Spannungen an Gate U_G , Drain U_D und Source U_S , die bezüglich Bulk-Potential angegeben werden.

Für *strong-inversion* / *forward saturation* ergibt sich die Transistorbeziehung zu:

$$I_D = \frac{\beta}{2n} \cdot [U_G - V_{T0} - n \cdot U_S]^2 \quad \text{für } \begin{matrix} U_S \leq V_P \\ U_D > V_P \end{matrix} \quad (\text{A.17})$$

mit der *pinch-off*-Spannung

$$V_p = \frac{U_G - V_{T0}}{n} \quad (\text{A.18})$$

Die Transistorkonstante $\beta = \mu \cdot C_{ox}' \cdot W_{eff}/L_{eff}$ bestimmt sich aus der Beweglichkeit der Ladungsträger μ , der flächenbezogenen Oxidkapazität C_{ox}' und der Transistorgeometrie W_{eff}/L_{eff} , n ist die Steilheitskonstante, V_{T0} ist die Schwellspannung.

Für den *weak-inversion*-Bereich lautet der Zusammenhang:

$$I_D = I_0 \cdot e^{\frac{U_G - V_{T0}}{nU_T}} \cdot \left(e^{\frac{U_S}{U_T}} - e^{\frac{U_D}{U_T}} \right) \quad (\text{A.19})$$

mit dem Sättigungsstrom $I_0 = 2 \cdot n \cdot \beta \cdot U_T^2$ und Temperaturspannung U_T .

Anhang B

Biologischer Anhang

Anhang B - Biologischer Anhang

1 Sehprozeß

1.1 Abbildung

Die Abbildung des Auges führt wie bei allen Loch- und Linsenkameras zur Zentralperspektive, die dadurch definiert ist, daß von jedem Punkt im Raum eine Linie durch den Knotenpunkt der Linse, das Projektionszentrum, gebildet werden kann. Der Schnittpunkt dieser Linie mit der Bildebene ist das Abbild des Raumpunktes (Bild B.1). Bei der Zentralperspektive werden Geraden im Raum als Halbgeraden abgebildet. Ausnahmen sind frontoparallele Geraden, die auf vollständige Geraden abgebildet werden und Geraden durch den Knotenpunkt, die auf Punkte abgebildet werden. Die Bilder paralleler Geraden sind im Abbild nicht parallel, sondern treffen sich in einem Fluchtpunkt. Ausnahme sind frontoparallele Geraden, die wieder auf parallele Geraden abgebildet werden. Zu jeder Geradenrichtung im Raum, mit Ausnahme der frontoparallelen, gibt es einen Fluchtpunkt. Dies ist ein Punkt im Bild, dem sich das Bild eines Objektes annähert, wenn es entlang dieser Geraden vom Betrachter weg wandert. Umgekehrt ist jeder Bildpunkt Fluchtpunkt einer bestimmten Geradenrichtung. Die Fluchtpunkte aller Geraden in einer Ebene bilden den Horizont dieser Ebene. Die Horizonte paralleler Ebenen sind gleich.

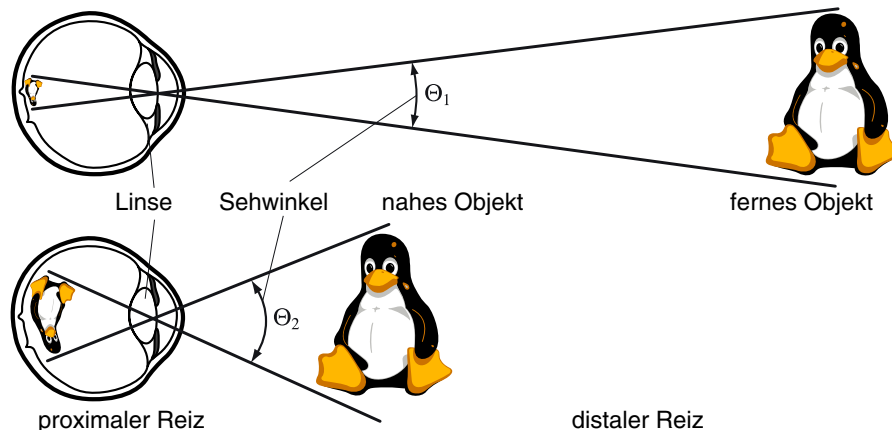


Bild B.1 Abbildung im Auge

Die Abbildungsfunktion des Auges ist sehr komplex, da ein mehrfacher Wechsel des optischen Mediums stattfindet (Luft, Hornhaut, Augenkammern, Linse und Glaskörper). Das Linsensystem des Auges kann jedoch vereinfacht als dicke Linse aufgefaßt werden und entsprechend wird die Abbildung durch die Kardinalpunkte Hauptebenen und Knotenpunkte beschrieben:

$$\frac{1}{b} = \frac{1}{g} + \frac{1}{f}$$

mit der Bildweite b , der Gegenstandsweite g und der Brennweite f .

Die Bildweite des Auges ist durch dessen Länge bestimmt und beträgt im Mittel 22,80 mm. Die Brennweite auf der Außenseite ist im Verhältnis der Brechungsindizes Luft und Kammerwasser auf 17,06 mm verkürzt. Nach der allgemeinen Abbildungsgleichung wird für näher am Auge befindliche Objekte eine größere Brechkraft benötigt, um eine scharfe Abbildung auf die Netzhaut zu erhalten. Durch den Linsenmuskel wird typisch eine Variationsbreite der Brechkraft der Linse von 15,8 dpt bis 25 dpt²⁷ erreicht, wobei die obere Grenze sehr stark vom Lebensalter abhängt: Kleinkinder bis 30 dpt, Erwachsene nur noch 20 dpt. Die Gesamtbrechkraft des Auges als Summe der Einzelbrechkraft beträgt 54,9...64,1 dpt²⁸.

1.2 Natürliche Abbildungsfehler

Das optische System Auge vermittelt eine vergleichsweise schlechte Abbildung der Umwelt auf die Netzhaut. Dies gilt vor allem für den Netzhautbereich außerhalb der Fovea. Die wichtigsten Ursachen für die natürlichen Abbildungsfehler sind:

- Sphärische Aberration

Bei einfachen optischen Linsen werden Strahlen am Rand vergleichsweise zu stark gebrochen als in ihrem Zentrum. Die Strahlen treffen sich dann nicht mehr in einem einzigen Brennpunkt. Eine kleine Pupille kann Abbildungsfehler durch sphärische Aberration vermindert, gleichzeitig erhöht sich die Schärfentiefe (Bild B.2). Unter Schärfentiefe versteht man dabei die Größe des Bereichs vor und hinter der Fixationsdistanz, in welchem Objekte noch scharf gesehen werden. Da mit kleinerer Pupille jedoch weniger Licht ins Auge dringt und sich damit die rezeptiven Felder vergrößern, wird die verbesserte Abbildung durch eine Verschlechterung der Erfassung begleitet. Das Auge kann jedoch mit Einstellung eines bestimmten Pupillendurchmessers an eine optimale Auflösung adaptieren.

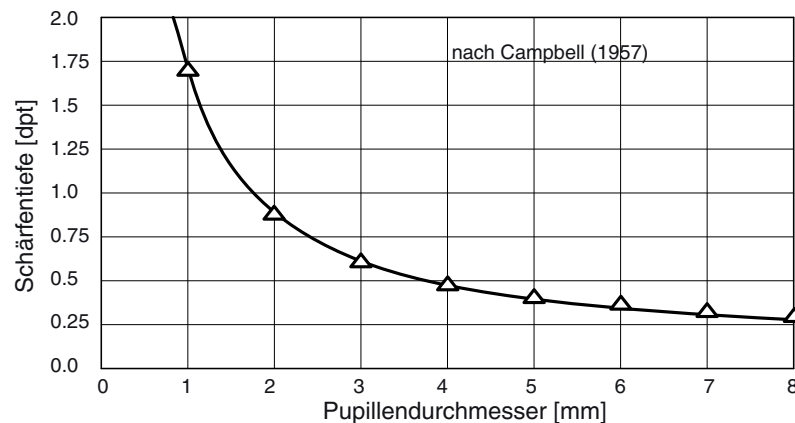


Bild B.2 Schärfentiefe in Abhängigkeit des Pupillendurchmessers

- Chromatische Aberration

Aufgrund der Dispersion wird Licht unterschiedlicher Wellenlänge von der Linse unterschiedlich stark gebrochen. Kurzwelliges Licht (Blau) wird dabei stärker gebrochen als langwelliges (Rot). Mehrfarbigen Gegenständen werden daher z.T. unscharf abgebildet. Akkommodiert das Auge z.B. auf die roten Linien eines mehrfarbigen Seobjekts, so erscheinen dessen blaue Linien unscharf und umgekehrt. Der maximale chromatische Fehler beträgt 1,0-1,5 dpt.

Die Sehschärfe hängt auch von der Farbe des empfangenen Lichtes ab. Sieht man einen Gegenstand bei monochromatischem Licht, so sprechen nur die Rezeptortypen an, die diese Wellenlänge vorwiegend als Reiz verarbeiten können. Die Dichte der aktivierten Rezeptoren ist damit geringer und mit ihr die Sehschärfe. Der Aufbau der rezeptiven Felder bedingt zudem, daß die Sehschärfe für reine Farbkontraste, welche keine Kontraste in der Leuchtdichte aufweisen, bis auf 30% (rot-grün oder blau-gelb) reduziert sein kann.

27. Als Maß für die Brechkraft einer abbildenden Optik wird die Dioptrie [dpt] verwendet. Brechwert D und Brennweite f sind umgekehrt proportional, es gilt $D = 1/f$ mit $1 \text{ dpt} = 1 \text{ m}^{-1}$. Ein normalsichtiges Auge besitzt bei Blick ins Unendliche per Definition eine Brechkraft von 0 dpt. Bei Blick in eine Distanz von 50 cm beträgt sie 2 dpt.

28. Die Brechungsindizes der am Abbildungsvorgang beteiligten Elemente sind: Luft $n=1$, Cornea $n=1.376$, Kammerwasser $n=1.336$, Linse $n=1.386$. Die Brechkräfte betragen: Cornea 42,8 dpt, Linse 15,8...25 dpt, konkave Rückseite der Cornea $-3,7 \text{ dpt}$, Gesamtbrechkraft des Auges $=54,9...64,1 \text{ dpt}$.

- Trübungen der Augenoptik und Fehlsichtigkeit

Da die Augenlinse aus Proteinlamellen aufgebaut ist, bewirken kleine Unregelmäßigkeiten im Aufbau eine Streuung des Lichtes, die sich besonders bei starkem Lichteinfall bemerkbar macht. Denselben Effekt bewirken gelöste Makromoleküle im Glaskörper. Das Streulicht überlagert sich dem Netzhautbild und man spricht von physiologischer Blendung. Mit der Trübungen des dioptrischen Apparates im Alter z.B. durch Einwanderung von roten Blutkörperchen (Erythrozyten), nimmt die Anzahl der Streuzentren zu. Das Gehirn gleicht die meisten Abbildungsmängel durch die zentrale Informationsverarbeitung aus. Diese Fähigkeiten der kortikalen Verarbeitung lassen sich besonders anschaulich an der Kompensation des blinden Flecks veranschaulichen. Zusätzlich können noch individuelle Abbildungsfehler auftreten. Kann beispielsweise ein Gegenstand vom dioptrischen Apparat nur unscharf auf der Retina abgebildet werden, so liegt Fehlsichtigkeit vor. Dabei unterscheidet man Kurzsichtigkeit (Myopie), Weitsichtigkeit (Hyperopie, Hypermetropie), Astigmatismus (zylindrische Fehlsichtigkeit, Stabsichtigkeit) und Alterssichtigkeit (Presbyopie). Beispielsweise ergeben 0.1 mm Abweichung der Augenlänge einen Brechkraftunterschied von 0.34 dpt ($1/0.017\text{m} = 58.82\text{ dpt}$; $1/0.0171\text{ m} = 58.48\text{ dpt}$) und entspricht einer Verschiebung des Fernpunktes von Unendlich auf eine Distanz von 2,9 m. Fehlsichtigkeit läßt sich durch zusätzliche Linsen im optischen Pfad korrigieren (Brillen, Kontaktlinsen). Weitere krankhafte Beeinträchtigungen des Sehapparates sind z.B. die Trübung der Linse (Katarakt, grauer Star), die medizinische Eingriffe erfordern.

1.3 Transduktion

Der Vorgang der Absorption eines Lichtquants bis zur Erzeugung einer Potentialänderung eines Rezeptors, der Hyperpolarisation, wird als Transduktion bezeichnet. Der Phototransduktionsprozeß ist die Umwandlung von Lichtreizen in chemische und elektrische Signale, die anschließend über den Sehnerv an das Gehirn weitergeleitet werden. Der Phototransduktionsprozeß wird durch verschiedene Proteine und Rückkopplungsmechanismen gesteuert.

Der Sehfärbstoff unterscheidet sich in den zwei verschiedenen Lichtrezeptorentypen des Auges. In den Stäbchen findet sich das Rhodopsin (auch Scotopsin). Dieses besteht aus dem Glycoprotein Opsin und Farbstoff 11-cis-Retinal, einem Aldehyd des Vitamin A. Von den Sehstoffen in den weniger empfindlichen Zapfen sind bisher das Jodopsin (besonders bei Säuger und Vögeln) und das Cyanopsin (bei Reptilien und Fische) bekannt, die ebenfalls einem Eiweißkörper und ein Vitamin-A-Derivat enthalten. Bei Wirbeltieren befinden sich die Sehpigmente in stapelförmigen Membranausstülpungen an der äußeren Zellmembran der Rezeptorzellen.

Bei Absorption eines Lichtquants zerfällt Rhodopsin durch photochemische Stereoisomeration in Opsin und all-trans-Retinal (Bild B.3). Das lichtaktivierte Retinal löst eine Enzymkaskade aus, indem Transducin, ein GTP²⁹-bindendes Protein aktiviert wird, das aus drei Untereinheiten

29. G-Proteine: Guaninnucleotid-bindende Proteine; GTP-Bindungsproteine oder GTP-bindende Proteine. Es handelt sich hierbei um eine große Superfamilie von Proteinen, welche die Fähigkeiten besitzen, GTP (Guanosin-5'-triphosphat) zu binden und dessen Hydrolyse in GDP (Guanosin-5'-diphosphat) und Pi (anorganisches Phosphat) zu katalysieren. Durch diesen Vorgang wird ein Schaltermechanismus für jeweils bestimmte Zellaktivitäten in Gang gesetzt. G-Proteine fungieren als Vermittler (Transducer) in der Signaltransduktion zwischen verschiedenen Rezeptoren (G-Protein gekoppelte Rezeptoren), die extrazelluläre Reize aufnehmen und den an diese Signalwege gekoppelten Effektoren, welche die Antwort in die Zelle weiterleiten.

(α , β , γ) besteht. Die α -Untereinheit löst sich dabei von β - und γ -Untereinheit, nachdem ein Austausch von GDP mit GTP stattgefunden hat und aktiviert das Enzym cGMP-Phosphodiesterase (cyclo-Guanosinmonophosphat). Dieser Vorgang dauert weniger als eine Picosekunde und die Quantenausbeute der Photoisomerisation beträgt ca. 70%.

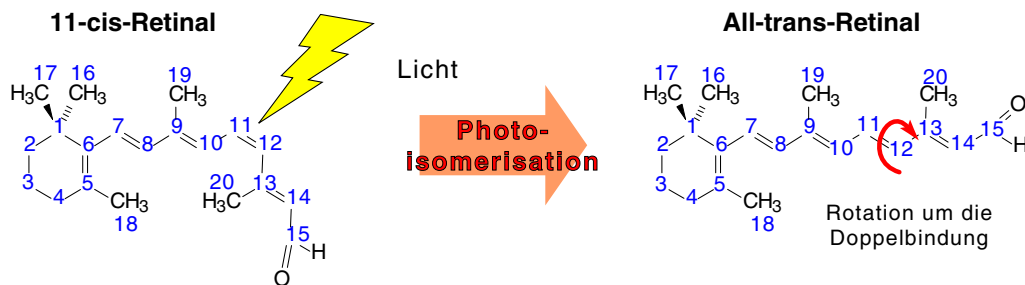


Bild B.3 Prinzip der Aktivierung von Retinal durch photochemische Isomerisation

Die cGMP-Phosphodiesterase beginnt nach der Aktivierung, das zyklische cGMP in 5'-GMP zu hydrolysieren. cGMP dient zur Steuerung der Kationenkanäle in der Zellmembran, indem es sich an ein spezifisches Protein bindet und die Kanäle öffnet. Damit diffundieren Calcium²⁺- sowie Kalium⁺-Ionen in die Zelle. Die Natrium-Kalium-Pumpe schleust hingegen Calcium und Kalium aus der Zelle und Natrium in die Zelle. Wird durch die cGMP-Phosphodiesterase die Konzentration von cGMP verringert, schließen sich die Kationenkanäle und Calcium²⁺- und Kalium⁺-Ionen können nicht mehr in die Zelle gelangen. Da Natrium-Kalium-Pumpe mit gleicher Aktivität weiterläuft, werden weiterhin Calcium²⁺-Ionen aus der Zelle geschleust und die Calciumkonzentration in der Zelle nimmt ab. Durch die entstehende Ladungsverschiebung kommt es zur Hyperpolarisation der Zelle. In Ruhe sind die Photorezeptoren mit ca. 30 mV depolarisiert. Bei Hyperpolarisation bricht diese Spannung um bis zu 25 mV ein. Über die angeschlossenen Bipolarzellen wird diese Erregung weitergeleitet.

Durch Absinken der intrazellulären Ca²⁺-Konzentration wird Recoverin freigesetzt, das die Guanylatcyclase aktiviert. Dabei wird neues cGMP aus GTP hergestellt, so daß die Kationenkanäle wieder geöffnet werden und ein Ionenausgleich stattfinden kann. Das ursprüngliche Rezeptorpotential wird wiederhergestellt, die Zelle depolarisiert. Bei Ansteigen der Ca²⁺-Konzentration in der Zelle bindet sich Ca²⁺ an Recoverin, so daß die Guanylatcyclase gestoppt wird. Calcium aktiviert ebenfalls die Phosphodiesterase bzw. hemmt die Rhodopsinkinase. Es hat somit auch Einfluß auf die Sensitivität der Photorezeptoren und die Adaptation an verschiedene Lichtverhältnisse. Guanylatzyklase resynthetisiert das Transducin, indem es die α -, β - und γ -Untereinheiten zu einem Trimer vereinigt. Rhodopsin wird entweder durch das Enzym Retinal-Isomerase aus all-trans-Isomerase oder aus der Vitamin A1 resynthetisiert. Durch Anlagerung, Abbau, Phosphorylierung und Dephosphorylierung aktivierender bzw. hemmender Substanzen, sowie durch die Konkurrenz zweier Proteine um die gleiche Bindungsstelle entsteht ein Gleichgewicht von Hemmung und Aktivierung. Dies ermöglicht die Steuerung der Zelle durch eine negative Rückkoppelung mit hoher Zeitkonstante und ein biochemisches Recycling-Prinzip. Diese Kombination von physikalischen und chemischen Prozessen ermöglicht eine hohe Effektivität und ist auch bei anderen chemischen Rezeptoren zu finden.

Eine wichtige Eigenschaft der Lichtsinneszelle ist die Reizverstärkung. Dies wird dadurch gewährleistet, daß ein einziges lichtaktiviertes Rhodopsinmolekül in einer ersten Stufe 500 Trans-

ducinmoleküle aktivieren kann und jede aktivierte cGMP-Diphosphoesterase vermag ca. 2000 Moleküle cGMP in 5'-GMP zu hydrolysieren. Der gesamte Vorgang besitzt somit eine Verstärkung von ca. 60dB und wird als Phototransduktionskaskade bezeichnet. Die Empfindlichkeit der Sehzellen höherer Wirbeltiere ist dabei so groß, daß nur ein einziges Lichtquant vom Retinal aufgenommen werden muß, um eine Erregung der Zelle auszulösen und mit 5 bis 14 Quanten ist bereits eine Empfindung im Gehirn möglich. Die Reaktionszeit des Rhodopsin liegt zwar im Pikosekundenbereich, die Latenz der Sehzelle ist jedoch sehr viel größer. So haben die Stäbchen beim Menschen eine Ansprechzeit von ca. 300 msec. Zapfen hingegen reagieren auf einen Lichtimpuls bereits nach 80 – 90 msec.

Für die Photorezeption ist das Retinal universell, da die Augen aller Tiergruppen sowie auch die verschiedenen Rezeptortypen das gleiche Grundmolekül verwenden. Auch die unterschiedlichen Absorptionsspektren sind nicht auf biochemisch verschiedene Photorezeptoren zurückzuführen. Die unterschiedlichen spektralen Eigenschaften sind durch verschiedene Anteile des Opsin am Rezeptor bedingt. Die Universalität des Retinals spricht bei aller Verschiedenheit des Augenbaus für eine ursprüngliche Homologie der Sehorgane. Diese Homologie ist u.a. durch die Identifizierung eines universellen Augen-Gens belegt worden [HCG95], [YR99].

Bei Primaten wird diskutiert, ob die Evolution der beiden Rezeptortypen im mittel- und langwelligen Bereich eng mit der Lebensweise im immergrünen Regenwald verbunden ist. Die genauere Wahrnehmung von Grün-, Gelb- und Rotabstufungen soll demnach der verbesserten Erkennung von essbaren Früchten und der Beurteilung ihres Reifegrades dienen [OV96]. Eine vergleichende Übersicht über das Farbsehen bei Säugetieren gibt [Jaco93].

1.4 Hell-/Dunkeladaptation

Diese Fähigkeit des menschlichen Auges, sich an unterschiedliche Lichtverhältnisse anzupassen, wird Hell-/Dunkeladaptation genannt. Das Auge vermag dabei, visuelle Informationen bei Tageslicht als auch bei geringer Helligkeit aufnehmen und verarbeiten. Diese Hell-/Dunkeladaptation beruht auf einer Änderung der Pupillenweite, der Adaptation der rezeptiven Felder und dem Wechsel von Zapfen- auf Stäbchensehen. Die Veränderung der Pupille kann die auf das Auge fallende Lichtintensität nur um den Faktor 25 regulieren. Um den vom Auge wahrnehmbaren Helligkeitsbereich von $1:10^{10}$ zu verarbeiten, genügen die zwei verschiedenen Rezeptorsysteme (Stäbchen- und Zapfensehen) nicht, sondern die Rezeptoren müssen zusätzlich an die jeweilige Leuchtdichte adaptieren. Dies erfolgt durch die Einstellung der Reizschwelle, indem sich in den Rezeptoren die Konzentration des Sehpigmentes mit der Beleuchtungsstärke ändert. Diese chemische Anpassung bei der Dunkeladaptation (Adaptation von hell nach dunkel) dauert vergleichsweise lange, da die Sehpigmente nur sehr langsam regenerieren. So passen sich zunächst die für das Farbsehen zuständigen Zapfen an das verringerte Umgebungslicht an. Nach etwa drei bis vier Minuten haben sie ihre maximale Empfindlichkeit erreicht. Die viel empfindlicheren Stäbchen adaptieren jedoch sehr viel langsamer an die Umgebungslichtstärke. So dauert es beim Übergang zum skotopischen Sehen eine bis mehrere Minuten, bis man Gegenstände überhaupt erkennt und ca. 40 Minuten bis die vollständige Adaptation eintritt (Bild B.4).

Beim Wechsel von einer dunklen in einer hellen Umgebung werden die Stäbchen schnell übersättigt und fallen zur Informationsgewinnung aus. Die Zapfen übernehmen diese Funktion und passen sich durch Einstellung eines biochemischen Gleichgewichts zwischen Regeneration und

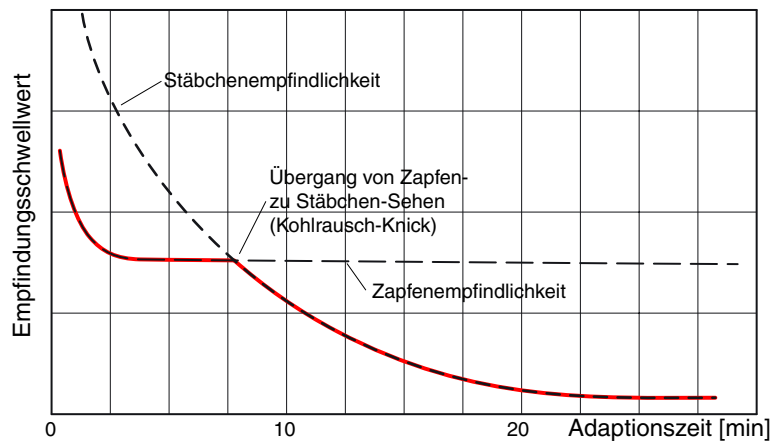


Bild B.4 Hell-Dunkel Adaptation des menschlichen Auges

Zerfall des Sehfärbstoffes an die herrschende Beleuchtungsstärke an. Dieser Vorgang wird als Licht- oder Helladaptation bezeichnet und benötigt nur ca. 5 Sekunden. Die subjektiv wahrgenommene Helligkeit ist dabei eine logarithmische Funktion der Lichtintensität.

Die neuronale Verarbeitung des Bildes in der Netzhaut ermöglicht einen weiteren schnellen Adaptationsmechanismus. Der erregende Teil der Netzhautfläche einer retinalen Ganglienzelle vergrößert sich mit zunehmender Dunkelheit. Der Nachteil der dadurch resultierenden geringeren Sehschärfe wird durch eine höhere Empfindlichkeit des Feldes ausgeglichen. Bei großer Helligkeit ist der erregende Teil der rezeptiven Felder klein. Neben der besseren Sehschärfe wird so auch eine übermäßige Reizung der Ganglienzellen verhindert [Wen03].

1.5 Weber-Fechner-Gesetz

Die Lichtrezeptoren des Auges können wie viele Sinnes- und Nervenzellen Reizintensitäten von mehreren Zehnerpotenzen Intensitätsunterschied verarbeiten. So besitzt weißes Papier im Vollmondlicht eine Leuchtdichte von ca. $10^{-2} \text{ cd} \cdot \text{m}^{-2}$, eine beleuchtete Straße ca. $10 \text{ cd} \cdot \text{m}^{-2}$ und der klare Mittagshimmel am Horizont kann ca. $10^5 \text{ cd} \cdot \text{m}^{-2}$ aufweisen. Die Funktion des Auges ist jedoch trotz der starken Intensitätsunterschiede gewährleistet.

Der geringste Reizunterschied, den das menschliche Auge eben noch wahrnehmen kann, wird als Unterschiedsschwelle bezeichnet und hängt vor allem vom Absolutwert des Reizes ab. Allgemein wächst die Unterschiedsschwelle mit der Reizintensität. Je höher die mittlere Lichtintensität ist, desto größer muß auch der Helligkeitsunterschied von zwei verschieden stark beleuchtete Flächen sein, damit diese unterschieden werden können. Die relative Unterschiedsschwelle bleibt dabei jedoch konstant. Dieser Zusammenhang heißt Gesetz von Weber³⁰. Unterstellt man, daß das Webersche Gesetz nicht nur für Unterschiedsschwellen, sondern generell anwendbar ist, kann es in das Gesetz von Fechner³¹ übergeführt werden. Die Wahrnehmungsintensität S ist demnach dem Logarithmus der Leuchtdichte L proportional:

$$S = K_1 \cdot \log(L) + K_2 \quad (\text{B.1})$$

Dieses Gesetz ist für das Auge allerdings nur im Bereich mittlerer Intensitäten von 10 und 10^5 Troland³² gültig. Bei hohen Umfeldleuchtdichten ab ca. 10^5 td wird aufgrund beginnender Blendung ein größerer Intensitätsunterschied benötigt. Bei niedrigen Leuchtdichten ist die rela-

30. Ernst Heinrich Weber (1795 - 1878)

31. Gustav Theodor Fechner (1801 - 1887)

tive Unterschiedsschwelle ebenfalls erhöht und deutet auf den Übergang vom Zapfen- zum Stäbchensehen hin. Eine ähnliche Beziehung wie das Weber-Fechner-Gesetz gibt die Potenzfunktion von Stevens wieder. Hier hängt die Wahrnehmungsintensität (*Lightness* L^*) mit der 3. Wurzel der Leuchtdichte L zusammen:

$$L^* = 116 \cdot \sqrt[3]{\frac{L}{L_a}} - 16 \quad (\text{B.2})$$

L_a ist dabei die Leuchtdichte, auf welche das Auge adaptiert ist. Diese Funktion wird heute häufiger verwendet als diejenige nach Fechner und findet sich u.a. in den CIE Standards wieder. Wird bei konstanter Adaptationsleuchtdichte L_a die Darbietungsdauer variiert, so ergibt sich, daß für sehr kurze Darbietungszeiten ($T < 20\text{ms}$) die Empfindlichkeit des Auges für Leuchtdichteunterschiede umgekehrt proportional zu T sinkt. Ab einer genügend langen Darbietungszeit, der abhängig von der Adaptationsleuchtdichte ist, tritt eine Sättigung ein und der minimal wahrnehmbare Leuchtdichteunterschied bleibt auch für größere Darbietungszeiten konstant. Die Tatsache, daß die Empfindlichkeit sowohl proportional zu den geometrischen Ausmaßen der Helligkeitsdifferenz als auch zur Darbietungsdauer ist, läßt auf örtliche und zeitliche Integrationseigenschaften des visuellen Systems schließen.

1.6 Signalauswertung und Weiterleitung in der Retina

Bereits in den ersten Neuronenschichten der Retina wird eine Verarbeitung der Bildinformation ausgeführt. Eine Zapfenzelle gibt ihre Information zunächst an mehrere Horizontal- und Bipolarzellen weiter. Eine Ganglienzelle erhält die verarbeitete Information von mehreren Bipolar- und Amakrinzellen und somit von mehreren Rezeptoren [Har38]. Diese Vereinigung von Informationen vieler Zellen wird Konvergenz genannt. Im Gegensatz dazu spricht man von Divergenz, wenn von einer Zelle Informationen zu vielen anderen Zellen fließen. Die Stäbchen (75-150 Million) und Zapfen (6.5 Million) konvergieren auf ca. 1.2 Million Ganglienzellen je Auge. Stäbchen zeigen folglich eine stärkere Konvergenz als Zapfen. Das Zusammenfassen vieler Rezeptoren führt zu einer stärkeren Aktivierung der zugehörigen Ganglienzelle und damit zu einer größeren Empfindlichkeit. Zapfen weisen dagegen eine geringere Konvergenz auf. Die Ganglienzellen lassen sich von einzelnen Zapfen erregen und ermöglichen damit eine höhere Auflösung und bessere Sehschärfe. Weiterhin überlappen sich die rezeptiven Felder mehrerer Ganglienzellen gegenseitig und besitzen unterschiedliche Größen. In der Fovea sind die Felder sehr klein, zur Peripherie hin werden sie ausgedehnter. Zudem variiert das Ausmaß der rezeptiven Felder adaptiv mit der Umgebungslichtstärke.

Die Wirkung der Rezeptoren im rezeptiven Feld einer Ganglienzelle kann sowohl erregend als auch hemmend auf dessen Pulsrate wirken. Dabei existieren 2 Arten der Anregung einer Ganglienzelle: Bei On-Center-Zellen führt verstärkter Lichteinfall im Zentrum des rezeptiven Feldes zur Erhöhung der Pulsrate und umgekehrt (Bild B.5). Off-Center-Zellen hingegen verringern die Pulsrate bei stärkerem Lichteinfall im Zentrum. Dies wird durch zwei verschiedene Typen von Bipolarzellen ermöglicht. On-Zellen geben die Polarisierung der Rezeptorzellen

32. Troland ist das Maß für die retinale Leuchtdichte. Die Pupillengröße variiert mit der Leuchtdichte und somit auch die Beleuchtungsstärke auf der Retina. Deshalb wird eine Größe verwendet, welche die Pupillenweite berücksichtigt (auch retinale Leuchtdichte genannt). Die allgemeine Beziehung lautet: $L[\text{td}] = L[\text{cd}\cdot\text{m}^{-2}] \cdot A_{\text{Pupille}}$. Dabei ist 1 td definiert als $1 \text{ cd}\cdot\text{m}^{-2}$ bei einer Pupillenfläche A von 1 mm^2 .

direkt weiter, während Off-Zellen die Polarisierung invertieren. Wenn die angeschlossene Rezeptorzelle hyperpolarisiert ist, depolarisiert die Off-Bipolarzelle und umgekehrt. Die an dieser Bipolarzelle angeschlossene Ganglienzelle vermindert demzufolge ihre Pulsrate bei steigender Lichtstärke und damit verbundener stärkerer Reizung des Rezeptors. Retinale Neuronen können somit nicht nur auf Licht, sondern auch auf Dunkelheit reagieren. Zusätzlich nehmen die Bipolarzellen mit einem adaptiven Verstärkungsfaktor Einfluß auf die Stärke des übertragenen Signals. Die Horizontalzellen haben weiterhin eine hemmende Wirkung auf die zugehörigen Bipolarzellen und sorgen damit für laterale Inhibition. Dadurch verstärken sie die durch Kontraste im Lichteinfall verursachten Polarisationsunterschiede benachbarter Rezeptorzellen.

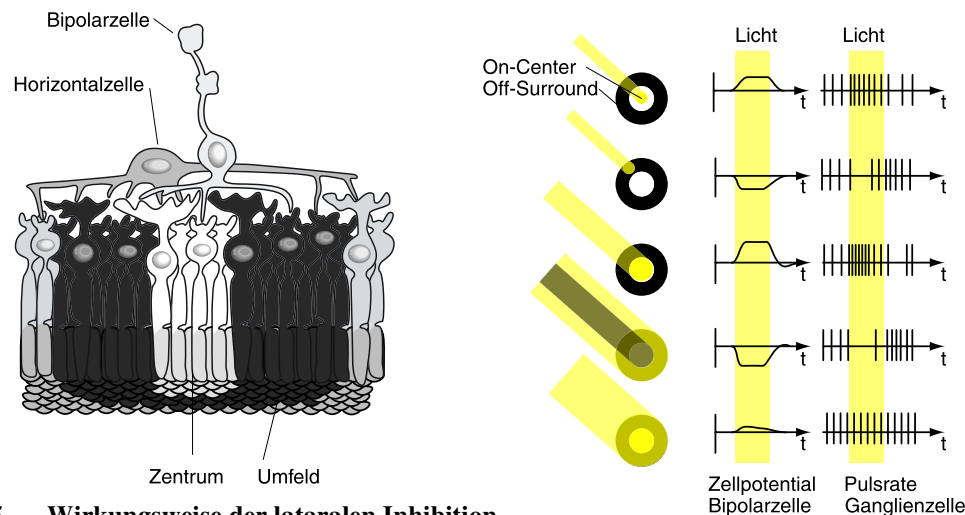


Bild B.5 Wirkungsweise der lateralen Inhibition

1.7 Kontrast, Kontrastverstärkung

Die Struktur der rezeptiven Felder führt zum Zentrum-Umfeld-Antagonismus und ist verantwortlich für den Effekt der simultanen Kontrastverstärkung. Darunter versteht man das Phänomen, daß von einer betrachteten Figur keine absolute Helligkeit, sondern Helligkeit in Abhängigkeit von Kanten wahrgenommen wird. Dies kann mit den sogenannten Mach-Bänder³³ (Bild B.6, links) veranschaulicht werden: Die Intensität der Kanten von kontrastierenden Flächen wird dabei verstärkt; eine helle Fläche wird an der Kante zu einer dunklen noch heller wahrgenommen, die dunkle Fläche an der Kante zur hellen noch dunkler. Ein anderes bekanntes Beispiel für den Simultankontrast stellt das Hermann-Gitter³⁴ dar (Bild B.6, rechts).

Diese Wirkungsweise der rezeptiven Felder zeigt, daß das Sehsystem darauf spezialisiert ist, eine adaptive Kodierung zur Übertragung von Kontrasten anstelle einfacher Absolutwerten von Helligkeiten vorzunehmen.

33. nach Ernst Mach (1838-1916)

34. nach Ludimar Hermann (1838-1914)

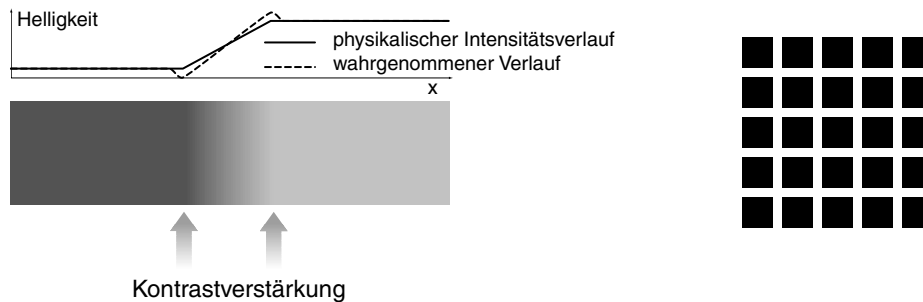


Bild B.6 Mach-Band (links) und Hermann-Gitter (rechts)

2 Kenngrößen des Auges

2.1 Sehschärfe

Die Sehschärfe ist allgemein die Fähigkeit, feine Strukturen auflösen und erkennen zu können. Sie wird im wesentlichen durch die Eigenschaften des optischen Apparates und der Größe der Rezeptoren bestimmt. Das menschliche Sehsystem ist jedoch in der Lage, auch Objekte zu registrieren, die in der Größe unter dem theoretischen Auflösungsvermögen der Rezeptorzellen liegen. Dieser Effekt wird als Überauflösung bezeichnet und wird durch die sakkadische Augenbewegung in Verbindung mit einer kortikalen Verarbeitung erreicht. Horizontal besitzt das menschliche Auge eine Auflösung von 3 bis 10 Bogensekunden bei einem Rezeptordurchmesser von ca. 30 Bogensekunden. Wesentlich schlechter ist jedoch die Wahrnehmung von Tiefeninformationen auf glatten Flächen. Bei einer sinusartigen Oberflächenstruktur erreicht die differenzierte Tiefenwahrnehmung nur etwa 4 Perioden pro Grad. Allgemein ist die Sehschärfe nach oben durch Beugungsphänomene begrenzt, die sich durch den minimal möglichen Pupillendurchmessers und die Zapfengröße ergeben.

Mit der formveränderlichen Linse vermag das Auge, Objekte in unterschiedliche Entfernungen scharf auf die Netzhaut abzubilden. Den entferntesten zu fokussierenden Punkt nennt man Fernpunkt und den naheliegendsten Nahpunkt. Idealerweise befindet sich der Fernpunkt in unendlicher Distanz. Ohne Fixationsobjekt wird die Akkommodationsruhelage eingenommen. Sie ist die am wenigsten beanspruchende Akkommodationseinstellung der Linse und befindet sich mit starken individuellen Unterschieden in 50 bis 100 cm Sehdistanz. Ein der Sehentfernung angepasster Akkommodationszustand ist Voraussetzung für eine scharfe Abbildung des Sehobjekts auf der Netzhaut. Jede Abweichung führt zu Unschärfe in der Abbildung.

Das Auflösungsvermögen des menschlichen Auges verhält sich im Bereich mittlerer Helligkeiten nahezu linear zur Lichtintensität, d.h. die Sehschärfe fällt proportional mit der Helligkeit ab (Bild B.7). Dieser Abfall der Sehschärfe bei mittleren Leuchtdichten wird durch die Variation der Größe der rezeptiven Felder in der Retina bewirkt. Mit abnehmender Helligkeit vergrößern sich die rezeptiven Felder einer Ganglienzelle. Um eine höhere Empfindlichkeit zu erreichen, werden die Reize einer größeren Anzahl von Rezeptorzellen zusammengefaßt. Dies wird jedoch durch eine Verminderung der Sehschärfe begleitet. Bei großer Helligkeit sind die rezeptiven Felder klein und sorgen für ein besseres Auflösungsvermögen. Gute Sehschärfe ist somit nur bei ausreichender Leuchtdichte zu gewährleisten. Jedoch verursachen zu große Intensitäten eine Blendung des Auges, womit die Sehschärfe nach einem Maximum bei ca. $10^3 \text{ cd}\cdot\text{m}^{-2}$ wieder abnimmt. Bei ca. $100\text{-}160 \text{ cd}\cdot\text{m}^{-2}$ wird optimale Sehschärfe erreicht. Diese mittlere Leuchtdichte

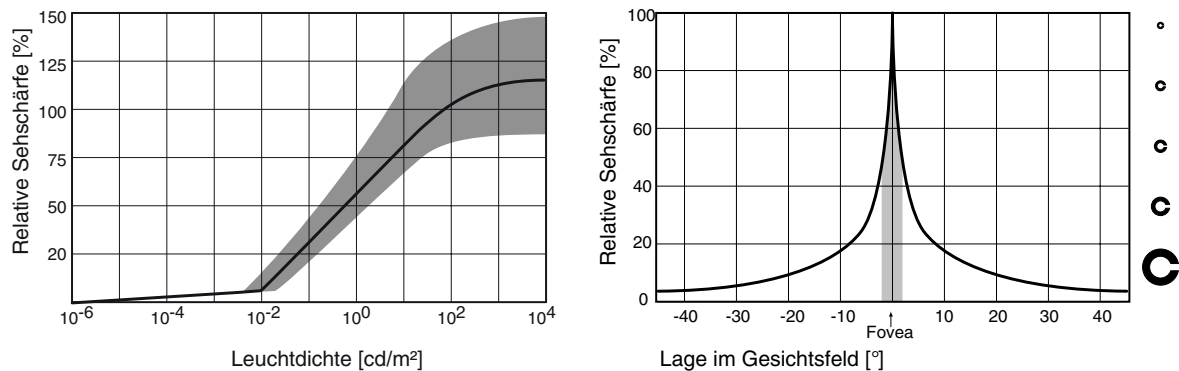


Bild B.7 Abhängigkeit der Sehschärfe von der Umgebungsleuchtdichte (links) (Streubereich für Beobachter im Alter zwischen 25 und 50 Jahren) und in Abhängigkeit von der Lage im Gesichtsfeld (beide [Schie01])

ist jedoch individuell und auch mit zunehmendem Alter verschieden [Schie01].

Da sich die Rezeptordichte und die Größe der rezeptiven Felder auch lokal auf der Netzhaut unterscheiden, hängt Sehschärfe auch vom Ort des Bildes auf Netzhaut ab. In der Fovea ist die Dichte der Zapfen am größten und zusätzlich sind auch die rezeptiven Felder klein, was eine gute Sehschärfe gewährleistet. Zur Peripherie hin nimmt die Anzahl der Zapfen pro Fläche stark ab und auch die Größe der rezeptiven Felder zu, wodurch eine schlechtere Sehschärfe bedingt ist (Bild B.8, Bild B.9). Diese Effekte führen dazu, daß nur ein fixierter Gegenstand scharf zu sehen ist, da er auf der Fovea abgebildet wird. Je entfernter das Bild von der Fovea ist, um so unschärfer erscheint ein Gegenstand. Die Sehschärfe fällt dabei exponentiell mit wachsender Exzentrizität ab (Bild B.7). Der maximale räumliche Sehwinkel bei ruhendem Auge beträgt ca. $\pm 80^\circ$ zur Augenachse. Dieses wahrnehmbare Gebiet wird als Gesichtsfeld bezeichnet. In der Fovea, bei einer Exzentrizität von 0° , ist die Sehschärfe am größten. Bereits bei ca. 30° Exzentrizität erreicht das Auge nur noch ca. 10% der maximalen Sehschärfe. Für ausreichende Sehschärfe relevant ist nur ein Bereich von weniger als 10° Sehwinkel. Die Fovea selbst hat funktionell nur ca. 5° Sehwinkel. Das entspricht einer Fläche mit einem Durchmesser von ca. 44 cm Durchmesser in 5 m Entfernung. Am Rand des Gesichtsfeldes werden Objekte und Bewegungen nur schemenhaft wahrgenommen. Durch Bewegen der Augen bei ruhiggestelltem Kopf kann ein Bereich von mehr als 180 Grad überstrichen werden, das Blickfeld

Die Sehschärfe ist auch vom Adaptationszustand des Auges abhängig. Im adaptierten Zustand ist schärferes Sehen möglich als bei nicht adaptiertem Auge. Wechselt man von einem hellen

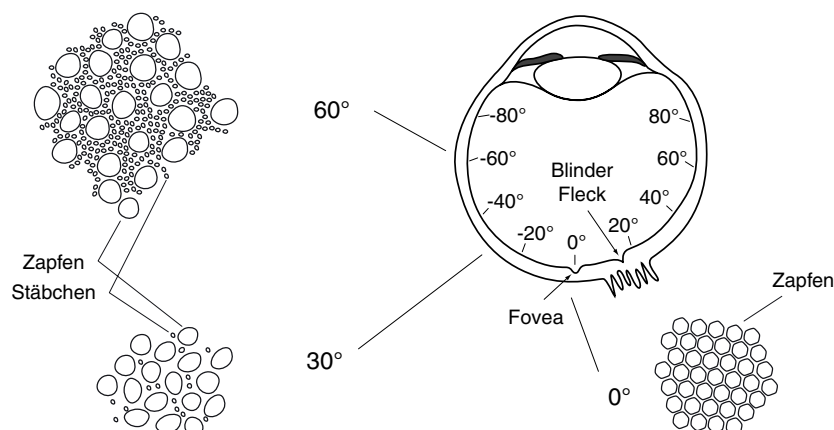


Bild B.8 Verteilung der Rezeptoren in Abhängigkeit von der Lage im Gesichtsfeld

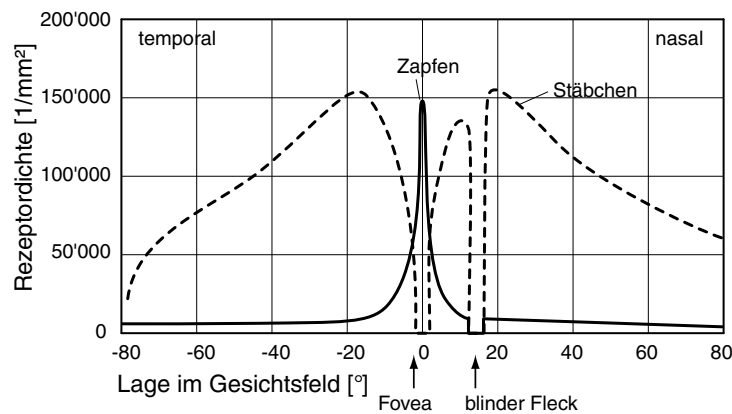


Bild B.9 Zapfen- bzw. Stäbchendichteverteilung bei einem seitlichen Schnitt durch das Auge [Hau94]

Raum in einen dunklen, so sieht man zunächst nichts, da die Zapfen nicht ausreichend erregt werden und die Stäbchen sich noch im Sättigungszustand befinden. Die Stäbchen erholen nur langsam, und je mehr Stäbchen in den Arbeitszustand übergegangen sind, desto deutlicher und schärfer wird das Objekt gesehen. Wechselt man umgekehrt von einem dunklen in einen hellen Raum, so werden sowohl die Stäbchen als auch die Zapfen zunächst übersteuert und es kann kein differenziertes Bild gesehen werden. Haben sich die Zapfen an die Grundhelligkeit angepasst, reagieren sie entsprechend der individuellen Reizintensität. Bei extrem schwacher Beleuchtung tragen nur noch die Stäbchen zur Wahrnehmung bei. Da die Stäbchen nicht wellenlängenselektiv sind, können zwar noch Formen wahrgenommen werden, aber keine Farben. Weiterhin können bei schwacher Beleuchtung die Objekte nicht mehr fixiert werden können, da in der Fovea keine Stäbchen vorkommen. Der Bereich des schärfsten Sehens ist somit nachtblind. Schwache Lichtpunkte in der Nacht sieht man demzufolge nur, wenn man knapp an ihnen vorbeisieht [Schm95].

2.2 Ortsfrequenzabhängigkeit

Zur Untersuchung der Empfindlichkeit des menschlichen Sehsystems auf bestimmte Ortsfrequenzen, können örtlich sinusförmige Muster verwendet werden (Bild B.10). Es zeigt sich, daß der Schwellwert der Wahrnehmbarkeit von der Ortsfrequenz abhängig ist. Bei hohen und tiefen Frequenzen bedarf es einer stärkeren Modulation, um eine Variation erkennen zu können. Diese Empfindlichkeit entspricht dem Betrag der örtlichen Modulationsübertragungsfunktion (MTF -

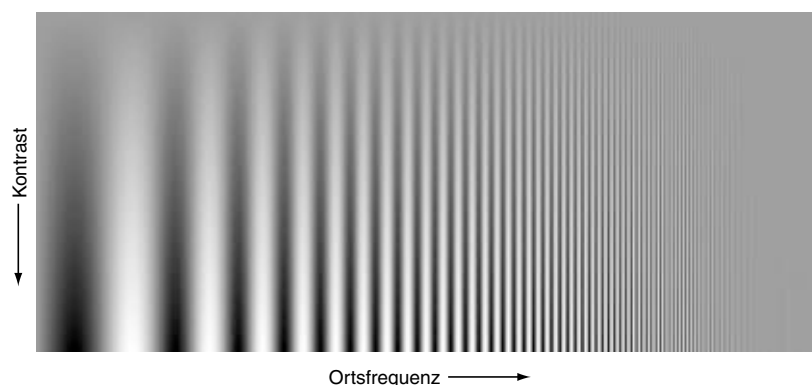


Bild B.10 Sinusgitter zur Veranschaulichung der Wahrnehmbarkeit verschiedener Ortsfrequenzen [Hau94]

Modulation Transfer Function). Die MTF besitzt eine Bandpaßcharakteristik, wobei der Abfall bei hohen Frequenzen stärker ausgeprägt ist als bei tiefen Frequenzen. Die Kontrastempfindlichkeit ist bei mittleren Ortsfrequenzen im Bereich von etwa 2-10 Perioden pro Grad Sehwinkel maximal. Bei gutem Kontrast können Ortsfrequenzen bis zu 50 Perioden pro Sehwinkelgrad unterschieden werden. Frequenzen ab 60 Perioden pro Grad Sehwinkel werden nicht mehr aufgelöst, sondern erscheinen als einheitlich graue Fläche. Die MTF ist zudem nicht isotrop und hat für Muster mit zweidimensionalen Ortsfrequenzen einen anderen Verlauf. Die Kontrastsensitivität für verschiedene Frequenzorientierungen zeigt, daß das menschliche Sehsystem besonders horizontale und vertikale Strukturen bevorzugt. Diagonale Verläufe registriert das Auge ca. 10% schlechter [CM87]. Diese richtungsabhängige Empfindlichkeit ist ein Anzeichen dafür, daß orientierungsselektive Strukturen zur Verarbeitung von Seheindrücken existieren.

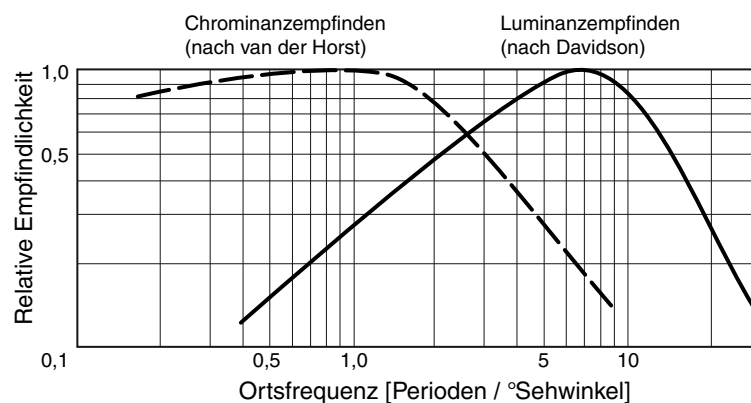


Bild B.11 Empfindlichkeit für Luminanz- und Chrominanzgitter in Abhängigkeit der Ortsfrequenz [Hau94]

In der Ortsfrequenzempfindlichkeit von Luminanz- und Chrominanzempfinden bestehen ebenfalls Unterschiede. Das Sensitivitätsmaximum der Modulationsübertragungsfunktion für Farbe liegt eine Größenordnung unter dem Maximum der Helligkeitswahrnehmung (Bild B.11). Das äußert sich darin, daß reine Farbkanten relativ schlecht wahrnehmbar sind. Bei der Erkennung von Strukturen und Objekten dominiert demnach die Luminanz. Bei Versuchen wurde jedoch festgestellt, daß Farbe bei der Wahrnehmung natürlicher Szenen hilfreich ist. Bildern mit Farbinformationen konnten dabei besser wiedererkannt werden. Bei Rot-Grün-Blinden konnte keine signifikanten Abweichungen festgestellt werden und führt zu der Vermutung, daß das Blau-Gelb-Farbsystem ausreicht, um natürliche Szenen einschätzen zu können.

Lange Zeit wurde eine getrennte Verarbeitung von Farb- und Bewegungsinformation im visuellen System postuliert, bei der die Verarbeitung von Bewegungsinformation allein dem Helligkeitsmechanismus zugeschrieben. Das Bewegungssehen galt als farbenblind und ist demzufolge unter isoluminanten Bedingungen unmöglich. Seit Beginn der 90iger Jahre gibt sich allerdings Berichte, daß auch ohne Helligkeitsinformation chromatische Bewegung wahrgenommen werden kann, wenn auch eingeschränkt [GP89], [Chic93], [SK95], [GH96].

Forschungen am magnozellulären System zeigen auch, daß deren rezeptive Felder aus Zentrum und Umfeld mit verschiedenen L- und M-Zapfenanteilen bestehen. Dadurch ist es nicht möglich, Isoluminanz über größere retinale Gebiete herzustellen. Das magnozelluläre Helligkeitssystem ist somit nicht völlig frei von Farbeinflüssen [Hunt71]. Diese Erkenntnisse sind u.a für die Entwicklung von Farbkodierungsverfahren für die Bewegtbildkompression wichtig.

2.3 Zeitliches Auflösungsvermögen

Mit der sogenannten Verschmelzungsfrequenz wird die Unterscheidungsschwelle kennzeichnet, ab der zeitlich aufeinanderfolgende Reize getrennt erfasst werden können. Sie ist stark von den äußeren Bedingungen abhängig und kann mehr als 90 Hz betragen. Bei Verringerung der Strahlungsintensität nimmt die wahrnehmbare Anzahl von Bildern pro Sekunde kontinuierlich ab (Bild B.12). Dieser Effekt wird u.a. bei der Projektion von Kinofilmen ausgenutzt, die in dunklen Räumen ausgestrahlt werden und so nur eine geringe Bildanzahl pro Sekunde benötigen. Beim Menschen verschmelzen in diesem Fall aufeinanderfolgende unbewegte Bilder,

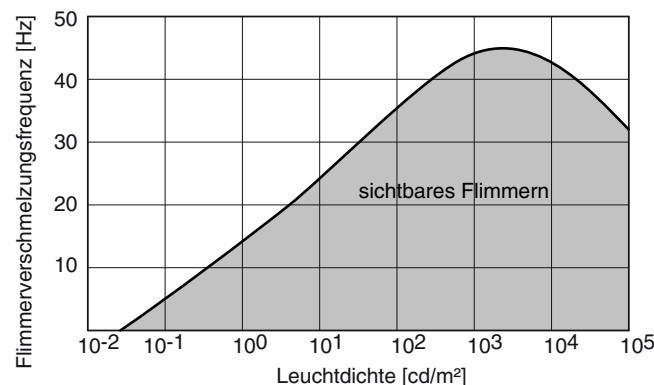


Bild B.12 Abhängigkeit der Flimmerfrequenz von der Leuchtdichte für foveales Sehen

wenn deren mehr als 20-25 pro Sekunde angeboten werden und die Darstellung erscheint flimmerfrei. Bei Frequenzen kleiner 20 Hz können jedoch bereits geringe Leuchtdichteschwankungen wahrgenommen werden. Des weiteren bewirken große flimmernde Felder eine höhere Verschmelzungsfrequenz als kleine [ISO9241]. Andere Lebewesen besitzen weit höheres zeitliches Auflösungsvermögen. So ist es beispielsweise Fliegen möglich, mehr als 240 Bilder pro Sekunde zu trennen. Dies wird allerdings durch ein wesentlich schlechteres örtliches Auflösungsvermögen begleitet. Im Gegensatz zur Sehschärfe ist das Bewegungssehen und das zeitliche Auflösungsvermögen im seitlichen Gesichtsfeld des Menschen aufgrund geringerer Rezeptordichte und höherer Nervenleitungsgeschwindigkeiten besser als in der Fovea (Bild B.13) [Krue93/1], [Krue93/2].

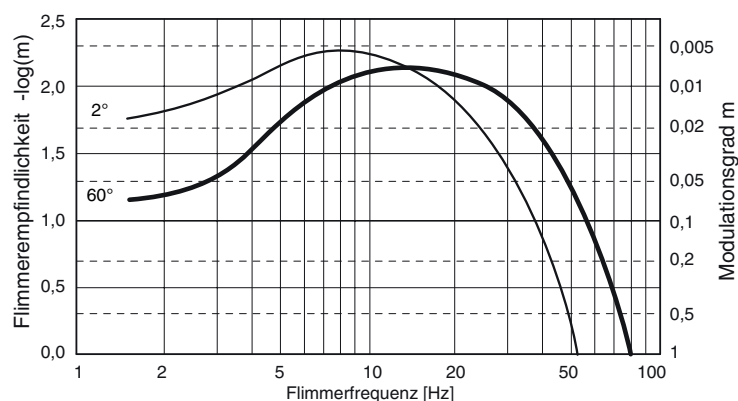


Bild B.13 Abhängigkeit der Flimmerempfindlichkeit von der Flimmerfrequenz für zwei verschiedene Sehfeldgrößen [Schie01]

Für das Bewegungssehen kann es bei schnell bewegten Objekten vorkommen, daß mehrere benachbarte Rezeptoren kurz hintereinander vom selben Punkt des Objekts gereizt werden, da zur Signalübertragung und -verarbeitung eine gewisse Zeit notwendig ist. Werden die Reize der beiden Rezeptoren praktisch gleichzeitig bearbeitet, so erscheint dies als lokale Unschärfe, die

als Bewegungsunschärfe bezeichnet wird. Wegen der unterschiedlichen Empfindlichkeit der Netzhaut für Bewegungswahrnehmung, können leichte Bewegungen mit der Peripherie der Netzhaut noch aufgelöst werden, während sie im direkten Blick eine Unschärfe verursachen. Die beim Menschen zu findenden Unterschiede in der zentralen und peripheren Auflösung lassen sich physiologisch auf die verschiedenen Ganglienzelltypen in der Retina zurückführen: Parvozellen sind neben der Farberkennung für hohe räumliche Auflösung bei niedriger zeitlicher Auflösung zuständig, während Magnozellen für die Helligkeitserkennung und die höhere zeitliche Auflösung verantwortlich sind. In der Retina sind Magnozellen hauptsächlich peripher vorhanden, während die Parvozellen hauptsächlich zentral angeordnet sind. Der Unterschied in der retinalen zeitlichen Auflösung des Menschen ist z.T. auch evolutionär als Fluchtreflex erklärbar: Gefahren kommen meist von der Seite. Durch die hohe zeitliche Auflösung in der Augenperipherie ist der Mensch in der Lage, dortige Änderungen schnell zu erfassen. Mit raschen Sakkaden der Augenbewegung werden diese Veränderungen mit dem hohen räumlichen Auflösungsvermögen der zentralen Retina, besonders in der Fovea centralis, genauer angesehen, um angemessen zu reagieren [Rock85].

Anhang C

Wahrnehmungstheorie und Farbbeschreibung

Anhang C - Wahrnehmungstheorie und Farbbeschreibung

1 Historie

1.1 Auge, Wahrnehmung und das Modell des Lichts

Die ersten Anfänge der Lehre vom Sehen stammen weder aus der Physik noch aus der Psychologie, sondern aus der Philosophie. Der Vorgang des Sehens war dabei eines der ersten Themen der Naturphilosophen. In der Antike erklärte Plato (427-347 v.u.Z.), daß Licht- bzw. Feuerstrahlen vom Auge ausgesendet werden und sich mit den von den Objekten ausgesendeten Repräsentationen vermischt. Epicurus (ca. 341-271 v.u.Z.) hingegen glaubt, daß Objekte Abbilder aussenden, die in das Auge gelangen und dort zu Wahrnehmung des Objektes führen. Das Konzept des Lichtes wurde somit ursprünglich entwickelt, um die Wahrnehmung von entfernten Objekten zu erklären. Die traditionelle Ansicht zur Physiologie des Auges stammt von dem arabischen Wissenschaftler Abu Ali Mohammed Ibn al Hazen (965-1039), wonach die Linse das wahrnehmende Organ darstellt. Leonardo da Vinci (1452-1519) und Johannes Kepler (1571-1630) erkannten dagegen die Ähnlichkeit des Auges mit einer Camera Obscura. Kepler folgerte, daß die Retina das wahrnehmende Organ sei, da durch die Linse ein spiegelverkehrtes Abbild auf der Retina zu sehen ist. Akkommodation erklärte er mit einer Verschiebung der Linse. Diesen Standpunkt vertrat auch René Descartes (1596-1650) aufgrund von Untersuchungen an einem Stierauge. 1688 beschrieb Edme Mariotte (ca. 1620-1684) den blinden Fleck als die Stelle, an welcher der optische Nerv aus der Retina austritt. Daraus schloß er, daß die Retina das Wahrnehmungsorgan sein mußte.

1.2 Physikalische Wahrnehmungs- und Farbtheorien

1.2.1 Von der Antike zur Neuzeit

Die Erkenntnisse zum Licht sind eng verknüpft mit dem Problem der Farben und der Farbtheorie. Eine Farbtheorie ist dabei eine meist spekulative Annahme über die Prozesse, die sich zwischen dem äußeren Lichtreiz und der Empfindung im Gehirn abspielen. So hielt man zum Beispiel in der Antike Farben für Mischungen von Licht und Dunkelheit. Aristoteles beschreibt beispielsweise sieben Farben, die sich zwischen den Extremen Schwarz und Weiß anordnen. Erst Leonardo da Vinci erbrachte wegweisende wissenschaftliche Erkenntnisse zur Farbtheorie. Das Universalgenie schuf ein Modell der Farbwahrnehmung und verwendete dabei drei Paare von Gegenfarben (Schwarz-Weiß, Rot-Grün, Blau-Gelb). Außerdem baute da Vinci ein Farbordnungssystem auf und benutzte Farbfilter, um Mischfarben zu untersuchen. Weiterhin entdeckte er das Phänomen des Simultankontrasts.

Existentielle Erkenntnisse auf der Grundlage experimenteller Beobachtungen zur Theorie der Farben stammen von Isaac Newton. Newton stellte fest, daß Farbe kein Zustand, sondern eine Sinneswahrnehmung ist („*The Rays to speak properly are not coloured. In them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Colour.*“). Die besondere Entdeckung von Newton war jedoch die spektrale Dispersion³⁵. Da der Brechungs-

35. Als Dispersion wird der Effekt bezeichnet, wenn Licht, das aus mehreren Wellenlängen besteht, aus einem optisch dünneren Medium in ein optisch dichteres Medium oder umgekehrt tritt und aufgrund unterschiedlicher Brechungsindizes für jede Wellenlänge verschieden stark gebrochen wird. Kurzwelliges Licht wird dabei stärker gebrochen als langwelliges.

der eine Funktion der Wellenlänge ist, ermöglicht diese Methode die Projektion eines kontinuierlichen Spektrums, wie es vom Regenbogen bekannt ist. Brechung wurde somit zu einem fundamentalen Prinzip der Farbanalyse. Newton demonstrierte die spektrale Dispersion durch die Zerlegung von weißem Sonnenlicht in die Spektralfarben mit Hilfe eines Prismas (Bild C.1). Durch eine Linse und ein zweites Prisma kann das zerlegte Licht wieder zu weißem Licht zusammengeführt werden. Ein weiteres Prisma ist jedoch nicht in der Lage, das durch das erste Prisma erzeugte Spektrum weiter zu zerlegen. Die Spektralfarben sind somit reine, nicht zusammengesetzte Farben. Aus seinen Beobachtungen schloß Newton, daß die Brechung des Lichts an den Linsen bei Teleskopen chromatische Aberrationen verursachen und entwickelte das Spiegelteleskop, das diese Fehlerquelle nicht aufwies.

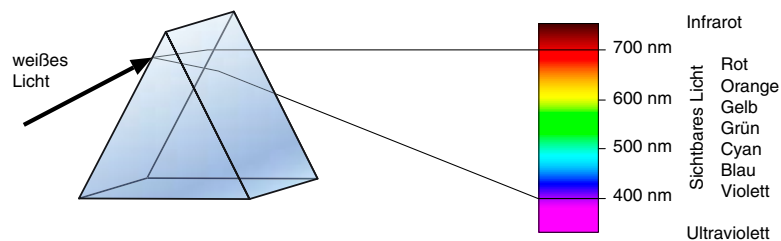


Bild C.1 Erzeugung eines Spektrums an einem Prisma

In Bezug auf die Farbanalyse nahm Newton fälschlicherweise an, daß es nur sieben homogene Farben gibt (Rot, Orange, Gelb, Grün, Cyan, Blau und Violett). Newton glaubte daher irrtümlich an die Ähnlichkeit zwischen Farben und den Tonintervallen einer Oktave. Die Theorie der Objektfarben, nach der die wahrnehmbare Farbe von Objekten von deren spektralen Reflektanz und somit von der selektiven Reflexion von homogenem Licht basiert, stammt ebenfalls von Newton. Weiterhin formulierte Newton zwei Gesetze der Farbmischung:

- 1. Farben lassen sich durch Komposition erzeugen, wobei gleiche Farben auf unterschiedliche Weise gemischt werden können. Aus je mehr Komponenten die Farben zusammengesetzt sind, desto entsättigter sind sie.
- 2. Unbunte Farben lassen sich aus bunten Farben ermischen; dazu werden nicht unbedingt alle Primärfarben benötigt. Außerdem lassen sich Farben mischen, die nicht gleich den reinen Farben sind.

1.2.2 Die Tristimulus-Theorie des Farbensehens

Die trichromatische Theorie des Farbensehens von Thomas Young (1773-1829) wurde 1802 veröffentlicht [Young1802]. Young betrachtet seine Theorie der Farbempfindung als eine Weiterentwicklung der Newtonschen Wellentheorie. Während Newton von nur sieben Grundfarben ausging, betrachtete Young eine Vielzahl verschiedener Wellenlängen aus dem Spektrum. Zu seiner Zeit war bereits klar, daß die Retina das eintreffende Licht in Nervenenerregung umwandelt. Young folgerte, daß auf der Retina ein physikalischer Mechanismus zur Farbwahrnehmung implementiert sein mußte, da ein einziger Punkt der Netzhaut unmöglich alle Bestandteile des Spektrums gleichzeitig an das Gehirn weiterleiten kann. Er postulierte daher im Jahre 1801 verschiedene retinale Substanzen, mit denen die Wellen des Lichts aufgenommen werden und über Nerven an das Gehirn geleitet werden. Youngs Theorie basiert auf den Newtonschen Erkenntnissen zur Farbmischung und der Beobachtung, daß alle wahrnehmbaren Farben aus drei Primärfarben gemischt werden können. Daraus schloß Young, daß drei verschiedene retinale Rezeptoren existieren, die Licht über einen begrenzten Wellenlängenbereich absorbieren und

empfindlich für Rot, Gelb und Blau sind. Später revidierte er seine Aussage und gab die Empfindlichkeiten mit Rot, Grün und Violett an. Da das gezeigte Violett jedoch vielen als Blau erschien, wurden allgemein die Farben Rot, Grün und Blau als Basisfarben der Tristimulus-Theorie verwendet.

1.2.3 Die Young-Helmholtzsche Theorie

Die Theorie von Young basiert auf der These, daß alle Farben aus drei Primärfarben gemischt werden können, obwohl es mehr als drei physikalisch verschiedene Lichtarten gibt. Wegen des Bezuges zur Wellentheorie des Lichts, die im Widerspruch zur der von Newton vertretenen Teilchentheorie des Lichtes stand, blieb die gesamte Theorie lange Zeit unbeachtet. Erst Hermann von Helmholtz (1821-1894) fand Bestätigung für die Annahmen in verschiedenen Farbmischexperimenten. Allerdings konnte Helmholtz nicht für jedes Farbpaar ein Mischverhältnis finden, das zu einer weißen Mischung führt. Anfangs gelang Helmholtz nur aus den Komplementärfarben Gelb und Blau eine weiße Mischung herzustellen. Hermann Günther Grassmann (1809- 1877) wies auf diese Unzulänglichkeit hin und Helmholtz konnte 1855 verbesserte Experimente zeigen. Im selben Jahr erbrachte auch James Clerk Maxwell (1831-1879) experimentelle Belege durch Farbmischungen, die den Vorhersagen der Youngschen Theorie entsprachen [Max1856]. Damit konnte die These, daß alle Farben aus drei einzelnen Primärfarben gemischt werden können, experimentell bestätigt werden. Die Wahrnehmung von Zwischenfarben wurde durch die unterschiedliche Erregung der drei Rezeptoren des Auges erklärt. Helmholtz stellte 1860 entsprechende spektrale Sensitivitäten vor [Helm24]. Maxwell machte diesen notwendigen Zwischenschritt explizit und bestimmte für sein Farbdreieck drei spektrale Empfindlichkeitskurven [Max1857]. Jedoch wußten sowohl Maxwell als auch Helmholtz, daß nicht für jede Farbe eine reales Spektrum zum Abgleich existiert. Sie wählten daher imaginäre Primärfarben für ihre Farbdreiecke (Bild C.2). Vom theoretischen Standpunkt ist bei diesen Überlegungen nicht klar, welche Reize tatsächlich als Primärfarben einzusetzen sind. Meist werden zwei davon aus den Enden des Spektrums ausgewählt. Zur Suche nach den physiologischen Primärfarben wurden später die Wahrnehmung von Trichromaten und Dichromaten verglichen. Die Ergebnisse liefern weitere Hinweise auf die Gültigkeit der Young-Helmholtzschen Theorie.

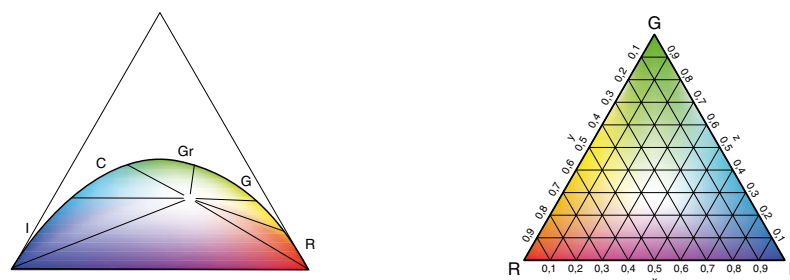


Bild C.2 Farbdreiecke von Helmholtz (links) und Maxwell (rechts)

Die trichromatische Theorie basiert somit auf der Beobachtung, daß Farbstimuli durch Mischung von anderen Stimuli vollständig nachgebildet werden können (siehe auch Kapitel II, Abschnitt 3.1). Dabei genügen jeweils drei Farbenkomponenten, um alle wahrnehmbaren Farbtöne zu mischen. Das menschliche Farbsehen folgt dem Prinzip der Superposition. Dies wurde zuerst von Grassmann erkannt, der die experimentellen Erkenntnisse in den sogenannten Grassmannschen Gesetzen zusammenfaßte:

- Durch die Mischung von drei geeigneten Primärfarben läßt sich jede gewünschte Farbe erzeugen.
- Für die Ermischung einer Farbe sind die Farbwerte der Primärfarben entscheidend, nicht ihre spektrale Zusammensetzung.
- Die Farbskala ist kontinuierlich.

Aufgrund von Linearität, Symmetrie und Transitivität der additiven Farbmischung ist jedes beliebige Set von drei unabhängigen Farbprimären zur Beschreibung von Farben geeignet. Die wichtigste Schlußfolgerung aus diesen experimentellen Entdeckungen war, daß das Auge über drei unabhängige Komponenten für die Farbwahrnehmung verfügen mußte. Moderne Darstellungen der trichromatischen Theorie geben Richter [Rich81], Wyszecki und Stiles [WS82] sowie Pokorny und Smith [PS86].

1.3 Phänomenologische Farbtheorien

1.3.1 Johann Wolfgang von Goethe

Im Gegensatz zu der physikalischen Modellierung der Wahrnehmung, die eher auf dem Experiment beruht, basiert die phänomenologische Betrachtung auf der oft spektakulären Erklärung von Phänomenen. Johann Wolfgang von Goethe (1749-1832) führte die phänomenologische Tradition an, die von Purkinje und Johannes Müller fortgeführt und später von Hering, Stumpf und Wertheimer vertreten wurde. Sie stellt die Gegenströmung zu dem Experimentalismus dar, der von Fechner und Helmholtz vertreten wurde. Das Wirken von Phänomenologen und Experimentalisten zeigt sich besonders auf dem Gebiet der visuellen Wahrnehmung.

Goethe, der mehr der Introspektion als der experimentell-empirischen Untersuchung vertraute, schrieb 1808 und 1810 seine zweibändige Farbenlehre. Der erste Teil dieses Werkes enthält die Beschreibung verschiedener subjektiver Phänomene wie Irradiation (das Ausbreiten von Weiß auf Schwarz), Hell- und Dunkeladaptation, positive und negative Nachbilder, Beleuchtungseffekte und Kontrastphänomene sowie Farbenblindheit. Farbsehstörungen waren zwar bereits im antiken Griechenland bekannt, jedoch erregte erst John Dalton³⁶ ein öffentliches Interesse an der Farbenblindheit. Dies führte letztlich zu der Unterscheidung zwischen der Farbe von Licht und der Farbe als Empfindung. Die im zweiten Band von Goethes Farbenlehre beschriebene Farbtheorie besteht hauptsächlich aus Polemik gegen Newton. Goethe wollte nicht glauben, daß weißes Licht aus verschiedenen Farben zusammengesetzt ist. Um Newton zu widerlegen, beschaffte er sich ein Prisma, hielt es gegen eine weiße Wand und fand, daß das hindurchgelassene Licht immer noch weiß aussieht, außer an den Rändern. Goethe behauptet, daß Farben als Mischung von Licht und Dunkelheit zu verstehen sind. Er kannte offenbar nur die subtraktive Farbmischung, die sich aus der Mischung von Farbpigmenten ergibt. Seine Farbtheorie selbst basiert auf einer Zwei-Farben-Hypothese mit den Grundfarben Gelb und Blau. Wegen der großen Berühmtheit von Goethe fand seine Farbenlehre dennoch längere Zeit Beachtung, während Youngs Farbtheorie weitgehend unbekannt blieb.

Johannes Evangelista Purkinje (1787-1869), Physiologe in Österreich, beschäftigte sich eben-

36. John Dalton (1766-1844), englischer Chemiker gilt als der Begründer der modernen Atomtheorie. 1794 beschrieb er seine eigenen Probleme, Chemikalien an ihrer Farbe zu erkennen. Daher rührt der frühere Begriff Daltonismus für Farbenfehlsichtigkeit.

falls mit den subjektiven Empfindungen in der Tradition von Goethe. So beschrieb er die Symptome, die sich bei Stimulierung des Auges durch elektrischen Strom und durch Ausüben von Druck ergeben. Er befaßte sich weiterhin mit den Phänomenen der Nachbilder, der Dunkeladaptation, dem blinden Fleck sowie Phänomenen des binokularen Sehens und der Farbempfindlichkeit in Abhängigkeit von der retinalen Exzentrizität. Besonders bekannt wurde das nach ihm benannte Purkinje-Phänomen, die Veränderung der Wahrnehmung von Farbton und Helligkeit unterschiedlicher Farben beim Wechsel von schwacher zu stärkerer Beleuchtung.

1.3.2 Die Duplizitätstheorie

Die Anatomie der Retina und die Beziehung zwischen Zapfen und Stäbchen wurden erstmals 1851 von Heinrich Müller (1820-1864) beschrieben, der zusammen mit Rudolph Albert von Kölliker (1817-1905) zu dem Schluß kam, daß zwei verschiedene Rezeptorarten auf der Retina angesiedelt sind. Müller entdeckte zudem das Sehpurpur in den Stäbchen. Max Schultze (1825-1874) zeigte 1866, daß die Stäbchen vor allem in der Netzhautperipherie auftreten und für das indirekte Sehen verantwortlich sind. Zapfen fehlen in der Retina von einigen nachtaktiven Tieren. Daher teilte Schultze den Stäbchen die Rolle des Nachtsehens zu und den Zapfen die des Tagessehens. Außerdem hielt er das Stäbchensehen für genetisch primitiver. Auch Charpentier, der die Befunde von Schultze nicht kannte, kam zu ähnlichen Ergebnissen. Aufgrund der Daten von Hermann Rudolph Aubert (1826-1892) und Purkinje postulierte er 1877 einen Helligkeitssinn und einen Farbsinn. Henri Parinaud (1844-1905) stellte 1881 eine ähnliche Theorie auf, die Theorie der zwei Retinas, in denen Zapfen und Stäbchen operieren.

Zu dieser Zeit wurde auch das Sehpurpur (Rhodopsin) eingehend untersucht. Franz Christian Boll (1849-1879) fand 1876 heraus, daß es unter Lichteinwirkung ausbleicht [Boll1877]. Dadurch konnte der Mechanismus der Dunkeladaptation und das Purkinje-Phänomen erklärt werden. König konnte schließlich die Absorptionsspektren von Sehpurpur (Rhodopsin) und Sehgelb (Jodopsin) bestimmen und zeigen, daß das Purkinje-Phänomen in den Unterschieden dieser Spektren begründet ist [Kön1894]. Johannes von Kries (1853-1928) integrierte diese Befunde zu seiner Duplizitätstheorie und postulierte 1894, daß das Dämmerungssehen von den Stäbchen vermittelt wird, während die Zapfen für das Tagessehen verantwortlich sind.

1.3.3 Die Heringsche Gegenfarbentheorie

Grundsätzlich stimmen alle Phänomenologen darin überein, daß es die vier Grundfarben Rot, Gelb, Grün und Blau, sowie Unbunt (Schwarz, Weiß) gibt. Ewald Hering (1834-1918) zeigte 1874, daß alle Farben zwischen den Grundfarben in einem Kreis angeordnet werden können (Bild C.3). Allerdings existieren jeweils Paare von sogenannten Gegenfarben, zwischen denen kein direkter Übergang möglich ist. Später vertrat Hering die Ansicht, daß Blau und Gelb die fundamentalen Farben sind, da die eine Hälfte seiner Farbordnung bläulich und die andere gelblich ist. Dabei wies Hering auf das Problem hin, das Gelb in der Dreifarbtheorie hatte. Helmholtz zufolge ergab sich Gelb als Mischung aus Rot und Grün, doch Hering postulierte, daß dies nicht mit der menschlichen Wahrnehmung übereinstimmt. Die Empfindung Gelb sei elementar und nicht auf eine Mischung zurückführbar. Hering argumentierte, daß niemals Mischungen von Rot und Grün auftreten, sie würden sich vielmehr auslöschen. Ein rötliches Grün ist nicht existent. Ernst Mach hatte dieses Problem bereits 1865 beschrieben und auch Aubert nannte

Fälle, in denen die Young-Helmholtzsche Theorie nicht zuzutreffen schien. So ergibt sich kein gelbliches Blau aus einer Mischung von Gelb und Blau, sondern Grün. Das Machsche Prinzip des psychophysischen Parallelismus, bei dem jedem Psychischen ein Physisches und umgekehrt entspricht, schien hiermit verletzt. Den Grundfarben mußte daher eine fundamentale Eigenschaft der physiologischen Prozesse zugrunde liegen.

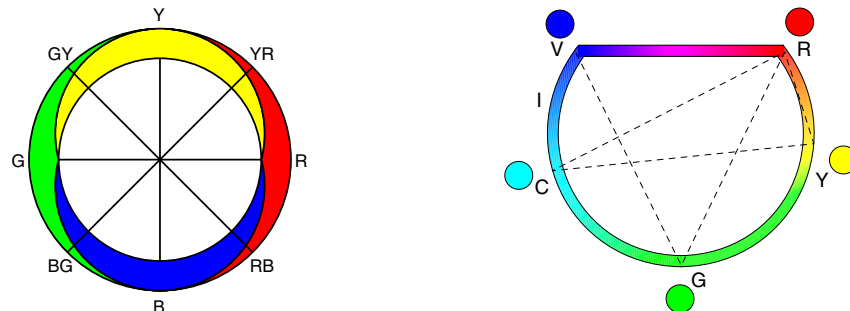


Bild C.3 Farbmolell von Hering (links) und Helmholtz (rechts)

1878 wurde von Hering die Gegenfarbentheorie postuliert [Her1878]. Dabei ging Hering von der Existenz von drei verschiedenen Substanzen in der Retina aus: einer Rot-Grün-, Blau-Gelb- und Hell-Dunkel-Substanz. Der Hell-Dunkel-Prozeß reagiert dabei auf Helligkeitskontraste, die Farbprozesse auf Farbunterschiede. Die Farbprozesse erregen zusätzlich auch den Schwarz-Weiß-Prozeß. Ein Problem der Heringsche Theorie lag jedoch darin, daß für das Schwarz-Weiß-System ein grundsätzlich anderer Mechanismus postuliert werden mußte. Diese Zusatzannahmen konnten von späteren Theorien nicht gehalten werden.

Damit standen sich Ende des 19. Jahrhunderts zwei scheinbar völlig gegensätzliche Theorien gegenüber: die Young-Helmholtzsche Tristimulustheorie und die Heringsche Gegenfarbentheorie. Beiden Theorien wurden durch Beobachtung der menschlichen Farbwahrnehmung gestützt. Helmholtz konnte seine Thesen mit den Ergebnissen von Farbmischexperimenten belegen und für Herings Ansatz sprachen verschiedene Wahrnehmungseffekte, wie beispielsweise Nachbilder, die in der Gegenfarbe des vorher gesehenen Bildes gesehen werden, der Einfluß der Farbintensität auf den Farbton bzw. Effekte wie Farbkonstanz, Farbumstimmung und der farbiges Simultankontrast oder die Tatsache, daß es möglich ist, grünliches Blau, nicht aber grünliches Rot oder gelbliches Blau wahrzunehmen.

Die Theorie der Komplementärfarben widerspricht jedoch nur auf den ersten Blick der Trichromatizitätstheorie. Physiologisch bestätigt werden konnten beide Theorien allerdings erst sehr viel später. Wie von Young vermutete und von Helmholtz aufgrund psychophysischer Experimente voraussagte, fand man in der Netzhaut drei verschiedene Farbrezeptoren, die jeweils im hoch, mittel und niederfrequenten Spektralbereich aktiv sind [SKB87].

Und auch die Gegenfarbentheorie konnte in Einklang mit der Trichromatizität bewiesen werden. Bei elektrophysiologischen Versuchen fand man die von Hering postulierten Gegenfarben als Eigenschaft der retinalen Ganglienzellen wieder. Die Parvo- und Koniozellulären Systeme setzen die Zapfenerregung in Gegenfarbeninformationen Rot-Grün und Blau-Gelb um und die rezeptiven Felder weisen aktivierende und hemmende Anteile auf. Hering erwartete zusätzlich zu den Gegenfarben noch einen Schwarz-Weiß-Prozeß. Diesem entspricht das Magnozellular System, zu dessen Erregung alle drei Zapfentypen beitragen. Herings Theorie kann somit durch die Verarbeitungsstufe der Netzhaut bestätigt werden. Zwei Lichtreize sind demnach verschie-

denfarbig, wenn die drei Zapfentypen in unterschiedlichen Anteilen stimuliert werden. Wenn die Stimulation des Magnozellulären Systems bei beiden Lichtreizen gleich ist, erkennt das Heligkeitsystem keinen Kontrast, die Farben sind isoluminant. Mit diesem Informationsverarbeitungsmodell lassen sich die Heringsche und die Helmholtzsche Anschauung zusammenführen.

1.3.4 Modifikationen der Farbtheorien

Am Ende des 19. Jahrhunderts wurden viele verschiedene Farbtheorien aufgestellt. Dabei wurden auch evolutionäre Elemente einbezogen. So geht die Evolutionstheorie der Farben von Schultze davon aus, daß die Stäbchen primitiver als die Zapfen sind und das Dämmerungssehen demzufolge primitiver als das Tagessehen ist. Von Hering stammt die Idee, daß die Rot-Grün-Substanz erst später entwickelt wurde als die Gelb-Blau-Substanz und daß die Schwarz-Weiß-Substanz die evolutionär älteste sei. Neben den evolutionstheoretischen Farbtheorien wurden auch solche entwickelt, die das zentrale Nervensystem berücksichtigen. Die Theorie von Selig Hecht (1892-1947) berücksichtigte nicht nur die Intensität der Reize und die dadurch ausgelösten photochemischen Reaktionen, sondern auch deren Dauer. Demzufolge zerfällt der Sehfärbstoff bei Lichteinfall und wird wieder regeneriert. In einem adaptierten Zustand liegt folglich ein charakteristisches Gleichgewicht zwischen Zerfall unter Einfluß des Reizes und Rekombination unter Einfluß der Konzentration der zerfallenen Substanz vor. Eine Weiterentwicklung dieser Theorie stammt von Rushton [Rush72].

1.4 CIE – *Commission Internationale d'Eclairage*

Im Jahr 1900 fand der *International Gas Congress* statt und führte zur Gründung der *International Commission on Photometry*, die sich mit den Fragen der Industrie jener Zeit aus dem Bereich der Photometrie leuchtender Gase (*incandescent gas mantles*) beschäftigte. Die folgenden Jahre brachten zahlreiche Fortschritte bei der Vermessung und Standardisierung der betreffenden Bereiche. Auch die Elektroindustrie zeigte zunehmendes Interesse an diesen Themen. Ab 1913 wurde daher eine größere Organisation geplant, die zur Gründung der *Commission Internationale d'Eclairage* (CIE, Internationale Beleuchtungskommission) führte. Die CIE ist heute eine technisch-wissenschaftliche ausgerichtete gemeinnützige Organisation, die sich selbst als Forum für die Themen versteht, die sich aus der wissenschaftlichen, technologischen und künstlerischen Betrachtung des Themas Licht und Beleuchtung ergeben. Der Schwerpunkt liegt dabei im Austausch von Informationen. Es werden grundlegende Standards und Prozesse für die metrische Erfassung des Themengebietes Licht und Beleuchtung entwickelt und Hilfe bei der Erarbeitung von nationalen und internationalen Normen geleistet. Seit 1999 beinhaltet Verantwortungsgebiet der CIE die optischen, metrischen und visuellen Aspekte der Kommunikation, Weiterverarbeitung und Reproduktion von Bildern, egal ob es sich dabei um analoge oder digitale Bildverarbeitungsmethoden, Speicher- oder Wiedergabemedien handelt. Die standardisierte Beschreibung der Farbwahrnehmung entstand unter Leitung der CIE Anfang des 20. Jahrhunderts durch empirische Versuche mit einer großen Anzahl normalsichtiger Personen zur Definition von Standard-Empfindlichkeitskurven. Dieser Standard repräsentiert seitdem das Farbsehen eines durchschnittlichen normalsichtigen Menschen. Trotz der Umstrittenheit dieser sogenannten Normalbeobachter, bilden die Versuchsergebnisse noch heute die Basis der modernen Farbmatrik.

2 Farbbeschreibungssysteme

Die Farbbeschreibung nach CIE-Standard stellt die Grundlage für alle praktisch relevanten Farbbeschreibungssysteme dar. Die technischen Spezifikationen basieren dabei direkt auf CIE XYZ. Ein System zur Klassifikation von Farben muß dabei in der Lage sein, jede Farbe mit hoher Präzision darzustellen. In der Wissenschaft der Kolorimetrie werden die Grundlagen für diese Beschreibungen untersucht. Die heute wichtigsten Farbraumbeschreibungen sind CIE XYZ, CIE xyY, CIE $L^*u^*v^*$ und CIE $L^*a^*b^*$.

Das Arbeitsgebiet der klassischen Kolorimetrie lag jedoch mehr in der Spezifikation von Farbe. Bei der Kodierung von Farbbildern treten andere Fragestellungen in den Vordergrund, die eine strenge Anwendung der Kolorimetrie unpassend erscheinen lassen. Ein Farbbild kann allgemein als eine Matrix von Bildpunkten aufgefaßt werden. Jeder Bildpunkt enthält dabei die numerische Definition der darzustellenden Farbe. Für die Auswahl eines geeigneten Systems zur Farbspezifikation muß beachtet werden, daß ein praktisches Bildkodierungssystem rechnerisch leistungsfähig sein muß und keine unbegrenzte Präzision aufbringen kann. Vielfach ist man auch der Ansicht, daß sich die Bildkodierung nicht vollständig an das CIE-Farbsystem anlehnen muß, da im allgemeinen nur ein Teil und nicht alle möglichen Farben kodiert werden müssen [Poy97]. Dies läßt sich jedoch nicht mit den Ergebnissen dieser Arbeit. Weiterhin müssen die Auswirkungen der Diskretisierungsfehler berücksichtigt werden. So werden in der Bildkodierung je nach Anwendung unterschiedliche Systeme zur Farbspezifikation benutzt. Praxisrelevante Farbbeschreibungssysteme zur Bildkodierung sind beispielsweise lineares RGB, nichtlineares RGB, nichtlineares CMY, nichtlineares CMYK und Ableitungen wie YUV oder Y'CbCr etc. Die im grafischen Bereich häufig verwendeten Größen Farbton und Sättigung sind dabei weder für eine präzise Farbbeschreibung noch für die Kodierung geeignet, können jedoch der praktischen Veranschaulichung von Farben dienen.

2.1 CIE XYZ und CIE xyY

Die Systeme CIE XYZ und xyY sind bei selbstleuchtenden Quellen universell anwendbar. Bei Farben, die durch Reflexion erzeugt werden, wie z.B. bei Fotografie, Druck oder Malerei, ist der resultierende Farbeindruck nicht nur eine Funktion des Farbmittels, sondern auch vom Leistungsdichtespektrum der Beleuchtung abhängig. Dies erschwert die objektive Beschreibung des Farbmittels, da eine Farbe nur für gegebene Randbedingungen spezifiziert werden kann. Das CIE XYZ-System definiert dabei die Überführung einer spektralen Leistungsdichtefunktion in drei numerischen Komponenten, die eine Beschreibung der jeweiligen Farbe wiedergeben. Diese Komponenten stellen die Basis für alle anderen Farbraumbeschreibungen dar. Der größte Nachteil des CIE XYZ-Farbraumes ist jedoch dessen Nichtgleichförmigkeit.

2.2 Lineares/nichtlineares RGB und CMY/CMYK

Während der CIE XYZ-Farbraum durch imaginäre Primärfarben und reale Spektralwertfunktionen gekennzeichnet ist, besitzt die Klasse der RGB-Farbräume reale Primärfarben. Dementsprechend sind die zugehörigen Spektralwertfunktionen imaginär. Das Koordinatensystem der linearen RGB-Räume wird dabei durch die Wahl der drei Grundfarben festgelegt. Im Prinzip sind alle RGB-Systeme gleichwertig, da sie durch Lineartransformation ineinander und in CIE XYZ überführbar sind. Da sich RGB-Systeme im allgemeinen auf positive ganzzahlige

Koeffizienten beschränken, ist nur der Teil des Farbraumes nutzbar, der sich innerhalb des von den Primärfarben aufgespannten Polygons befindet. Bei der Umrechnung verschiedener RGB-Räume treten allgemein Verluste auf, da nur die Farben erhalten bleiben, die in beiden Farbräumen existent sind. Die Diskretisierung führt weiterhin zu Rundungsfehlern bei der Transformation und zu einer schlechteren Auflösung besonders im Blaubereich aufgrund der kleineren MacAdam-Schwellenellipsen in diesem Bereich. Nichtlineare RGB-Systeme erweitern die Spezifikation eine durch eine Gammakorrektur, mit der das Verhalten von Kathodenstrahlröhren bei Monitoren beschrieben wird. Nichtlineares RGB besitzt zwar eine bessere Gleichförmigkeit als CIE XYZ bei vergleichsweise geringem Rechenaufwand, ist jedoch nicht zur Spezifikation von Farbabständen geeignet.

Der entscheidende Vorteil der RGB-Systeme insbesondere gegenüber CIE XYZ ist, daß sie eine direkte Beschreibung von linearen, additiv arbeitenden Anzeigegeräten darstellen. Am Gerät selbst ist somit keinerlei Umrechnung der Farbbeschreibung notwendig. Die von Videostandards spezifizierten RGB-Systeme orientieren sich zwar an den Eigenschaften von realen Monitoren, die physikalischen Vorrichtungen zur additiven Reproduktion von Farbe können jedoch entsprechend der technischen Realisierung große Unterschiede aufweisen. Als Austauschformat hat sich daher *Rec. 709* durchgesetzt, das eine standardisierte Verarbeitung im Studio, die Speicherung und Ausstrahlung bis zur Wiedergabe auf einem Monitor oder Projektor sicherstellt. Zur Spezifikation von Farbaufnahmesystemen sind RGB-Beschreibungen aufgrund der unrealen Spektralwertfunktionen jedoch völlig ungeeignet.

Während RGB-Beschreibungen auf der additiven Farbmischung (siehe auch Kapitel II, Abschnitt 3.1) und der Angabe von roten, grünen und blauen Farbanteilen beruhen, arbeiten CMY-Systeme mit der subtraktiven Farbmischung der Komponenten Cyan (C), Magenta (M) und Gelb (Y) (Bild C.4). Da die Mischung von praktischen Tinten jedoch kein reines Schwarz, sondern eher ein dunkles Braun ergibt, läßt sich die Farbgröße und Kontrastfähigkeit bei CMYK durch Zugabe von Schwarz (Black) als Druckbestandteil verbessern. Der Farbumfang von CMYK-Systemen ist jedoch aufgrund der Begrenzungen von Farbprimären und Farbmischung wesentlich kleiner als bei RGB.

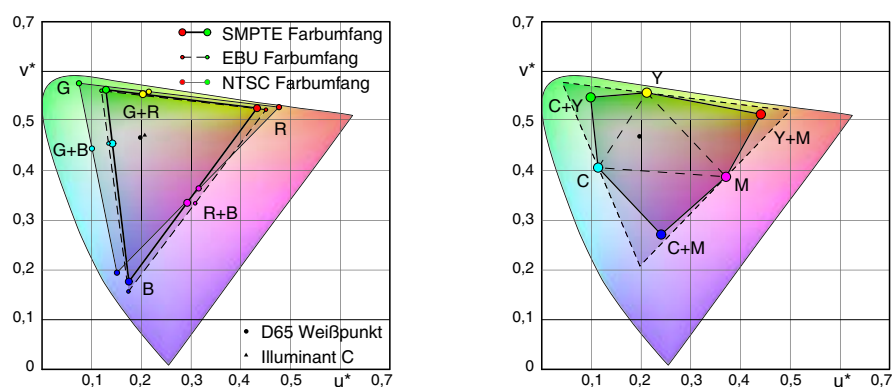


Bild C.4 Farbumfang verschiedener RGB-Systeme (links) und eines CMY-Systems (rechts) im CIE L*u*v-Diagramm

2.3 CIE L*a*b

Als Standard zur Farbbeschreibung von nicht selbstleuchtenden Körpern, wie Textilien, Mal-farben, etc. wurde 1976 das CIE L*a*b-System eingeführt. Dabei handelt es sich um einen drei-

dimensionalen Farbraum, der isoluminante Farben in Ebenen mit den im rechten Winkel zueinander stehenden Farbkoordinaten a^* und b^* anordnet. Senkrecht zu den isoluminanten Ebenen steht die achromatische Achse L^* , die von $L^*=0$ für Schwarz bis $L^*=100$ für Weiß definiert ist (Bild C.5). Die Berechnung der Koordinaten des CIE $L^*a^*b^*$ -Systems erfolgt durch Transformation von CIE XYZ auf Basis der Gegenfarbentheorie von Hering in Anlehnung an die Transformation der Rezeptorsignale im Auge. In Analogie zu den Heringschen Farbdifferenzen Rot-Grün und Blau-Gelb werden aus den Farbmaßzahlen X , Y und Z die Differenzen: $a' = X - Y$ und $b' = Y - Z$ gebildet. Daraus ergibt sich eine Skala, auf der positive Werte von a' eine Dominanz von Grün anzeigen, während bei negativen Werten von a' Rot überwiegt. Analog dazu ist die Koordinate b' im Blaubereich negativ und im Gelbbereich positiv. Die Helligkeit L^* einer Farbe wird weiterhin durch die Koordinate Y bestimmt. Da die XYZ-Werte jedoch bezüglich der Farbempfindung nicht gleichabständig sind, trifft dies auch für die daraus gebildeten Differenzen a' und b' zu. Zur Anpassung an die Nichtlinearität der Farbwahrnehmung wird entsprechend der Potenzfunktion von Stevens die 3. Wurzel der CIE XYZ-Koordinaten verwendet (Anhang B, Abschnitt 1.5).

$$\begin{aligned} L^* &= 116 \cdot f\left(\frac{Y}{Y_{\max}}\right) - 16 \\ a^* &= 500 \cdot \left(f\left(\frac{X}{X_{\max}}\right) - f\left(\frac{Y}{Y_{\max}}\right)\right) \\ b^* &= 200 \cdot \left(f\left(\frac{Y}{Y_{\max}}\right) - f\left(\frac{Z}{Z_{\max}}\right)\right) \end{aligned} \quad (C.1)$$

$$\begin{aligned} \text{mit} \quad f(x) &= \sqrt[3]{x} & \text{für } x > 0,008856 \\ f(x) &= 7,787 \cdot x + 0,138 & \text{für } x \leq 0,008856 \end{aligned}$$

Durch diese nichtlineare Transformation ergibt sich eine deutlich verbesserte Gleichabständigkeit der Farben in Bezug auf die Wahrnehmung. Im Vergleich verbessert sich die Gleichabständigkeit von ca. 20:1 bei XYZ auf ca. 3:2 bei CIE $L^*a^*b^*$. Nachteilig ist jedoch der vergleichsweise hohe Rechenaufwand. Im CIE $L^*a^*b^*$ -Farbraum ist im Gegensatz zum CIE L^*u^*v keine Sättigung definiert. Dies ist dadurch begründet, daß im CIE $L^*a^*b^*$ -System keine Farbtabelle angegeben ist, die eine Darstellung von Farbton und Sättigung erlaubt. Folglich ist keine sinnvolle Darstellung des Spektralfarbenzuges im CIE $L^*a^*b^*$ -Farbraum möglich. Die Spektralfarben können zwar im a^*b^* -Diagramm dargestellt werden, bilden aber nicht die äußere Begrenzung und nehmen daher keine bevorzugte Lage im Diagramm ein. Dies ist dadurch zu erklären, daß die Buntheit und nicht die Sättigung zur Darstellung gelangt [Schäp93]. Damit ist eine Messung der absoluten Sättigung in Relation zur Spektralkurve nicht möglich.

2.4 CIE L^*u^*v

Der CIE L^*u^*v -Farbraum (u und v sind hierbei nicht zu verwechseln mit U und V aus dem Videofarbraum YUV) wurde ebenfalls 1976 als Modifikation des 1964 auf Basis der CIE-UCS-Farbtabelle vorgeschlagenen Farbraumes eingeführt. Auch bei der Berechnung der Koordinaten von CIE L^*u^*v wird von den Farbmaßzahlen X , Y und Z ausgegangen. Die Berechnung bezieht Zwischenwerte u' und v' ein, die aus dem obsoleten 1960 CIE uv -System

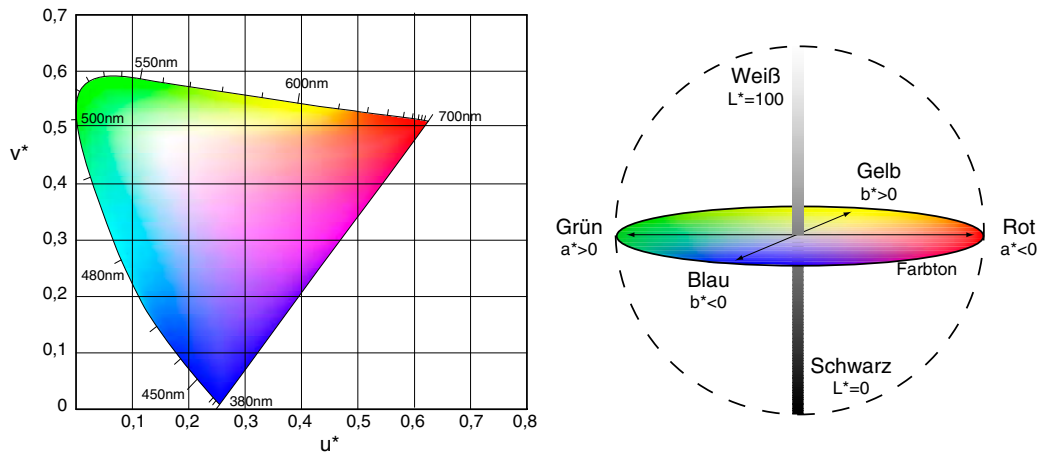


Bild C.5 Darstellungen des CIE L*u*v*- und CIE L*a*b*-Farbraumes

hervorgingen. Weiterhin sind die Koordinaten u'_n und v'_n für ein Referenzweiß notwendig.

$$\begin{aligned}
 L^* &= 116 \cdot f\left(\frac{Y}{Y_n}\right) - 16 \\
 u^* &= 13 \cdot L^* \cdot (u' - u'_n) \\
 v^* &= 13 \cdot L^* \cdot (v' - v'_n)
 \end{aligned} \tag{C.2}$$

$$\text{mit} \quad u' = \frac{4X}{X + 15Y + 3Z} \quad v' = \frac{4Y}{X + 15Y + 3Z}$$

Y_n, u'_n, v'_n = Werte für das Bezugsweiß

Die Berechnung von u' und v' erfolgt mit zwei linearen Gleichungen, L^* definiert wiederum die Helligkeit des Farbwertes nach der Potenzfunktion von Stevens. Die lineare Transformation hat den Vorteil, daß additive Mischungen auch in der $u'v'$ -Farbtafel auf einer Verbindungsgeraden liegen (Bild C.5). Die MacAdam-Ellipsen werden dabei allerdings nicht vollständig als Kreise wiedergegeben. Der Unterschied der Ellipsen verringert sich von 20:1 in der Normfarbtafel auf ca. 4:1. Die CIE L*u*v*-Farbtafel eignet sich zur Veranschaulichung von farbmtrischen Zusammenhängen der additiven Farbmischung, wie beispielsweise die Darstellung des Farbumfanges im Farbfernsehen. Ein weiterer Vorzug der $u'v'$ -Farbtafel gegenüber CIE L*a*b* ist die Tatsache, daß der Spektralfarbenzug als Begrenzung des real existierenden Farbraumes dargestellt werden kann. Damit ist es möglich, Farben in ihrer Relation zur maximal erzielbaren Farbsättigung darzustellen. Farbsättigung und Sättigungsdifferenz sind somit angebbare. Der größte Nachteil ist jedoch die nur bedingte Gleichabständigkeit, hervorgerufen durch die Linearität der Transformation. Das CIE L*u*v*-System wird vor allem zur Erfassung von Farbumterschieden bei selbstleuchtenden Körpern, z.B. bei Bildschirmen (DIN 6169 Teil 6), verwendet. In der Drucktechnik konnten sich L*u*v* und L*a*b* erfolgreich bewähren. Der notwendige Berechnungsaufwand ist allerdings zu groß, um in Echtzeit bei Bildschirmen Anwendung zu finden. Obgleich es nicht spezifisch für diesen Zweck optimiert wurde, bietet nichtlineares RGB für diese Anwendung ausreichende Uniformität bei angemessener Performance.

2.5 CIE-LCH

CIE-LCH ist kein neuer Farbraum im eigentlichen Sinne. Es handelt sich dabei lediglich um die Transformation von CIE $L^*a^*b^*$ (oder CIE $L^*u^*v^*$) in Polarkoordinaten. Dabei werden anstelle der kartesischen Koordinaten a^* und b^* die Polarkoordinaten $C^*_{ab} = \sqrt{a^{*2} + b^{*2}}$ und $h^*_{ab} = \arctan(b^*/a^*)$ verwendet (Bild C.6). Die Buntheit C^*_{ab} gibt den Abstand von der Unbuntachse und der Bunttonwinkel h^*_{ab} den Winkel an. Die Helligkeit L^* bleibt unverändert. Die Buntheit sollte dabei nicht mit der Sättigung verwechselt werden. Wird die Helligkeit einer Farbvalenz verändert, die Koordinaten a^* und b^* und damit auch die Größe C^*_{ab} aber konstant belassen, bewegt sich der Farbpunkt auf einer Geraden parallel zur Unbuntgeraden und nicht auf einer Geraden durch den Nullpunkt des Koordinatensystems, wie es bei konstanter Sättigung der Fall ist.

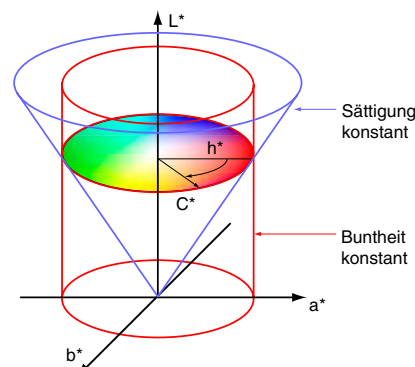


Bild C.6 Darstellung des CIE LCH-Farbraumes

2.6 Farbabstand ΔE

Mit den Koordinaten der Farbräume $L^*a^*b^*$ und $L^*u^*v^*$ lassen sich Farbabstände zwischen zwei Farbvalenzen berechnen:

$$\Delta L^* = L^*_1 - L^*_2$$

$$\Delta a^* = a^*_1 - a^*_2$$

$$\Delta b^* = b^*_1 - b^*_2$$

$$\Delta E = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (C.3)$$

Dabei gibt ein Farbabstand von $\Delta E=1$ einen Farbunterschied an, der von einem 2°-Normalbeobachter gerade noch wahrgenommen werden kann.

Farbabstand	Wahrnehmbarer Farbunterschied
$0 < \Delta E < 1$	Normalerweise nicht sichtbare Abweichung
$1 < \Delta E < 2$	Sehr geringe, nur von einem geschulten Auge erkennbare Abweichung
$2 < \Delta E < 3,5$	Mittlere, auch von einem ungeschulten Auge erkennbare Abweichung
$3,5 < \Delta E < 5$	Deutliche Abweichung
$5 < \Delta E$	Starke Abweichung

Allerdings reicht oft die Angabe des Gesamtfarbabstandes ΔE nicht aus. Dann ist es notwendig, die Einzelabstände mit anzugeben, um die Art der Abweichung näher zu beschreiben. Zwei Farben ergeben im $L^*a^*b^*$ -Raum zudem einen anderen Abstand ΔE als im $L^*u^*v^*$ -Raum. Mit dem Farbabstand muß daher auch der verwendete Bezugsfarbraum angegeben werden.

2.7 Gegenfarbsysteme

Neben dem $L^*a^*b^*$ -Farbraum existieren noch andere Farbbeschreibungen, die auf der Theorie der Gegenfarben von Hering basieren. Demnach gibt es drei gegensätzliche Farbpaare: Rot-Grün, Gelb-Blau und Hell-Dunkel. Im Farbdreieck liegen sich die Gegenfarben nicht genau gegenüber. Es ist jedoch möglich, die maximal gesättigten Farben auf einem Kreis anzuordnen, in dem sich die Gegenfarben gegenüberliegen. Dieser Kreis ist zwischen Blau und Rot durch die Farbwahrnehmung Purpur geschlossen. Das HSB-System der Farbmatrik lehnt sich an diesen Farbenkreis an. Die Darstellung der Größen Buntton (*Hue*), Sättigung (*Saturation*) und Helligkeit (*Brightness*) eignen sich dabei besonders zur intuitiven Veranschaulichung von Farben. Dabei entspricht die Helligkeit die Länge des Farbvektors, während Farbton und Sättigung den Winkel und den Radius in einer Polarkoordinatendarstellung mit dem Weißpunkt als Zentrum wie bei CIE-LCH beschreiben. Darstellungen in Polarkoordinaten werden dabei häufig aus Koordinaten von RGB-Systemen berechnet. Beispiele sind die Farbräume HSV und HSI. Für die Anwendung gilt zu beachten, daß sich die Auflösung und somit auch die Diskretisierungsfehler zwischen den Darstellungen in kartesischen und Polarkoordinaten unterscheiden. Umrechnungen zwischen verschiedenen diskreten Farbbeschreibungen führen somit zu Verfälschungen, die in unterschiedlicher Weise auffällig werden können (Bild C.7).

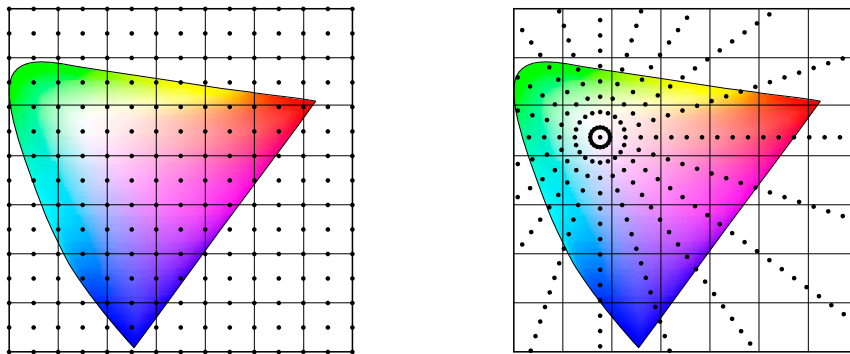


Bild C.7 Verschiedene Fehler durch Diskretisierung in kartesischen und Polarkoordinaten

2.8 Farbsysteme und Farbordnungssysteme

Es existieren zwei Möglichkeiten, einen gleichabständigen Farbraum zu ermitteln: Entweder wird eine einfache Umrechnung als oberstes Ziel angestrebt und die Gleichabständigkeit diesem Ziel untergeordnet oder es wird die bestmögliche Gleichabständigkeit gesucht und Probleme bei der Umrechnung in Kauf genommen. Ein Farbsystem (z.B. CIE $L^*a^*b^*$, CIE $L^*u^*v^*$) ermöglicht eine relativ einfache Transformation einer Farbvalenz aus den Ausgangswerten XYZ und die Berechnung von Farbabständen mit Hilfe euklidischer Geometrie. Da jedoch jeder Farbort in Abhängigkeit von Winkel und Abstand definiert sein muß und diese nicht vollständig mit der menschlichen Wahrnehmung korrespondieren, weisen Farbsysteme Defizite im Bezug auf die Gleichabständigkeit auf. Im Gegensatz dazu ordnen Farbordnungssysteme (z.B. Munsell, DIN, OSA, NCS, Coloroid, RAL-Design) die Farbvalenzen willkürlich an, wobei keine mathematische oder geometrische Definition der Farbe möglich ist. Die meisten Farbordnungssysteme systematisieren nach Helligkeit, Farbton und Sättigung. Systeme wie Munsell bauen auf Farbtafeln auf, die einen Vergleich von einzelnen Farben ermöglichen. So ist es z.B. möglich, die Farbe einer Tinte zu spezifizieren, indem die Anteile an Standard- (oder geheimen) Tinten angegeben werden, die gemischt werden müssen, um die entsprechende Farbe nachzu-

bilden. Die Funktionsweise von PANTONE basiert auf diesem Prinzip.

Das *Book of Color* von Munsell, das älteste Farbbordnungssystem, wurde erstmals 1915 von Albert H. Munsell (1858-1918) herausgegeben. Es basiert auf einer Ordnung der Farben nach dem Farbton in einem Farbkreis mit 5 Hauptfarben (Bild C.8). Das Munsell Book of Color umfaßt 40 Blätter, auf denen für je einen bestimmten Farbton die verfügbaren Helligkeits- und Buntheitsstufen veranschaulicht werden. Eine Farbbezeichnung in diesem System könnte zum Beispiel 5YR4/6 lauten. Dabei handelt es sich um ein rötliches Gelb, das zwischen Rot (10R) und Gelbrot (10YR) liegt, mit der Helligkeit 4 von 10 und der Buntheit 6 von 16 möglichen Stufen. Die visuelle Gleichabständigkeit des Systems ist dabei nur bei einem Betrachtungslicht der Lichtart C gewährleistet. Das schwedische *Natural Colour System* (NCS) von Anders Hård verwendet als Farbraumkoordinaten nicht wie üblich Farbton und Sättigung, sondern den Grün/Rot- und Gelb/Blau-Anteil, gemäß der Gegenfarbentheorie von Hering. Jede Farbe kann daher mit Ihrem Anteil an Rot/Grün, Blau/Gelb und Schwarz/Weiß bewertet werden. Der Farbkreis im NCS ist in vier Quadranten eingeteilt, die sich aus den Komplementärfarben ergeben (Bild C.8). Die Quadranten sind ihrerseits in 100 Farbtöne unterteilt. Der Schwarz/Weiß-Anteil ist nicht im Farbkreis dargestellt, es gilt jedoch die Vereinbarung, daß die Summe von Schwarz-, Weiß- und Buntanteil immer 100 beträgt. Eine Farbbezeichnung in NCS lautet beispielsweise 2030 G20Y, dabei handelt es sich um eine Farbe mit 20% Schwarz, 30% Weiß und einem Buntverhältnis von 80 Teilen Grün zu 20 Teilen Gelb. In der Gesamtmischung entspricht dies 40% Grün und 10% Gelb. Problematisch sind sowohl beim Munsell-System als auch beim NCS, daß die verwendeten Farbbezeichnungen sprachlich verwirrend sind. So werden dieselben Namen wie in physikalisch begründeten Farbsystemen verwendet, obwohl die Farben nicht übereinstimmen. So ist im Munsell-System die zu Rot komplementäre Farbe zwar als Blaugrün (BG) bezeichnet, das physikalische Blaugrün (Cyan) ist jedoch mit Blau (B) angegeben. Im NCS System wird die Komplementärfarbe von Gelb (Y) zwar korrekt mit Blau (B) benannt, jedoch wie bei Munsell farblich als Cyan dargestellt. Und komplementär zu Rot (R) ist bei NCS Grün (G).

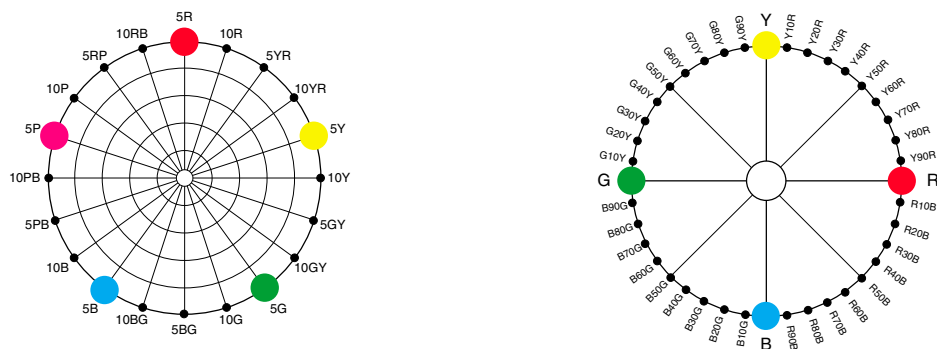


Bild C.8 Farbbordnung nach Munsell (links) und NCS (rechts)

Wie beim Munsell-System ist auch beim NCS-System keine direkte formelmäßige Umrechnung der Farbzahlen in CIE-Farbwerte möglich. Immerhin kann man bei beiden Systemen mittels Tabellen die CIE-Werte XYZ annähernd ermitteln. Farbbordnungssysteme haben zwar die Freiheit, mehr Rücksicht auf eine exakte Gleichabständigkeit der Farben zu nehmen, ohne mathematische Definitionen einhalten zu müssen. Sie sind jedoch aus diesem Grund zur Beschreibung von physischen Farben für Bilddaten u.a. ungeeignet. Der absolute Farbraum, also die Darstellung einer physiologischen Gleichabständigkeit von Farbe, gepaart mit einer einfachen Berechnung der Farbvalenzen und Farbabständen bleibt letztlich ein theoretisches Konstrukt.

Bemerkungen zu den japanischen Sprichworten

Die zu Beginn jedes Kapitels zitierten Zeilen sind japanische Sprichworte, die im wörtlichen Sinn eine Referenz zu Farben haben. Die Zuordnung der Sprichworte zu den einzelnen Kapiteln erfolgte willkürlich und steht in keiner Beziehung zu den im jeweiligen Kapitel gemachten Aussagen. Zur näheren Erklärung der Bedeutung sollen die Sprichworte kurz erläutert werden. Als lateinische Umschrift wird die direkte Hepburn-Umschrift ohne Lautersetzung verwendet.

しゅ まじ あか
朱に交われば赤くなる。(Shu ni majiwareba akakunaru) bedeutet wörtlich: „(Wer sich) mit Zinnober verbündet, wird rot.“ und entspricht übertragen in etwa dem Sprichwort „Wer mit Hunden zu Bett geht, steht mit Flöhen auf.“

しゅ うば むらさき
朱を奪う紫。(Shu wo ubau Murasaki), wörtlich: „Vom Zinnober gestohlenen Purpur.“, hat übertragen eine etwas komplizierte Bedeutung in der Art von „mit fremden Federn schmücken“. Purpur ist als Mischfarbe nach japanischer Auffassung unrein, während Zinnober als traditionelle Farbe für Shintou-Schreine als rein gilt. „Vom Zinnober gestohlenen Purpur“ ist somit der Versuch, einer unreinen Farbe, den Anschein von Reinheit zu geben.

やなぎ みどり はな くれなゐ
柳は緑 花は紅。(Yanagi ha midori, Hana ha kurenai) bedeutet wörtlich: „Die Wiese ist grün, die Blume ist rot.“ und steht als Beschreibung für Natürlichkeit im Kontrast zu künstlicher Nachahmung.

しろ くろ
白黒はつきりさせる。(Shiro Kuro hakkiri saseru), wörtlich: „Weiß und Schwarz werden getrennt gemacht.“ hat übertragen die Bedeutung, daß z.B. gute und schlechte Taten bzw. Schuld und Unschuld unvereinbar sind.

あおい あい で あい あおし
青は藍より出でて藍より青し。(Aoi ha Ai yori dedete Ai yori aoshi), bedeutet wörtlich: „Das Blau, das vom Indigo kommt, ist blauer als Indigo.“ und geht auf die japanische Färbetradition zurück. Die traditionelle Gewinnung des blauen Farbstoffes durch Fermentierung der Indigo-Pflanze ist ein sehr aufwendiger Prozeß, der eine lange Ausbildung erfordert. Um ein guter Färber zu werden, muß man zudem bei einem guten Meister in die Lehre gehen und wenn man sich anstrengt, wird man letztlich besseres Indigo herstellen können, als sein Meister. Das Sprichwort wird daher übertragen verwendet zur Umschreibung von z.B. einem Sohn, der mehr erreicht, als sein Vater.

